
Detection and correction of deformations in UAVs' images using deep learning

Leandro Henrique Furtado Pinto Silva



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Leandro Henrique Furtado Pinto Silva

**Detection and correction of deformations in
UAVs' images using deep learning**

Dissertação de mestrado apresentada ao Programa de Pós-graduação da Faculdade de Computação da Universidade Federal de Uberlândia como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Área de concentração: Ciência da Computação

Orientador: Dr. André Ricardo Backes

Coorientador: Dr. Mauricio Cunha Escarpinati

Uberlândia

2021

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

S586 2021	Silva, Leandro Henrique Furtado Pinto, 1991- Detection and correction of deformations in UAVs' images using deep learning [recurso eletrônico] / Leandro Henrique Furtado Pinto Silva. - 2021. Orientador: André Ricardo Backes. Coorientador: Mauricio Cunha Escarpinati. Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Ciência da Computação. Modo de acesso: Internet. Disponível em: http://doi.org/10.14393/ufu.di.2021.249 Inclui bibliografia. Inclui ilustrações. 1. Computação. I. Backes, André Ricardo, 1981-, (Orient.). II. Escarpinati, Mauricio Cunha, 1976-, (Coorient.). III. Universidade Federal de Uberlândia. Pós-graduação em Ciência da Computação. IV. Título. CDU: 681.3
--------------	---

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
 Coordenação do Programa de Pós-Graduação em Ciência da Computação
 Av. João Naves de Ávila, nº 2121, Bloco 1A, Sala 243 - Bairro Santa Mônica, Uberlândia-MG, CEP 38400-902
 Telefone: (34) 3239-4470 - www.ppgco.facom.ufu.br - cpqfacom@ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciência da Computação				
Defesa de:	Mestrado Acadêmico, 07/2021, PPGCO				
Data:	05 de maio de 2021	Hora de início:	09:00	Hora de encerramento:	11:30
Matrícula do Discente:	11922CCP009				
Nome do Discente	Leandro Henrique Furtado Pinto Silva				
Título do Trabalho:	Detection and correction of deformations in UAVs' images using deep learning				
Área de concentração:	Ciência da Computação				
Linha de pesquisa:	Ciência de Dados				
Projeto de Pesquisa de vinculação:	-				

Reuniu-se, por videoconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciência da Computação, assim composta: Professores Doutores: Bruno Augusto Nassif Travençolo - FACOM/UFU; Dalcimar Casanova - UTFPR; Mauricio Cunha Escarpinati - FACOM/UFU (coorientador) e André Ricardo Backes - FACOM/UFU, orientador do candidato.

Os examinadores participaram desde as seguintes localidades: Dalcimar Casanova - Pato Branco/PR; Bruno Augusto Nassif Travençolo, Mauricio Cunha Escarpinati e André Ricardo Backes - Uberlândia/MG. O discente participou da cidade de Uberlândia/MG.

Iniciando os trabalhos o presidente da mesa, Prof. Dr. André Ricardo Backes, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor presidente concedeu a palavra, pela ordem sucessivamente, aos examinadores, que passaram a arguir o candidato. Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o candidato:

Aprovado.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Mestre.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Mauricio Cunha Escarpinati, Presidente**, em 05/05/2021, às 11:44, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Bruno Augusto Nassif Travençolo, Professor(a) do Magistério Superior**, em 05/05/2021, às 11:45, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **André Ricardo Backes, Professor(a) do Magistério Superior**, em 05/05/2021, às 11:45, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Dalcimar Casanova, Usuário Externo**, em 05/05/2021, às 13:49, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2743075** e o código CRC **59C4780A**.

*To my mother,
and my grandmother (in memoriam).*

Acknowledgements

To God, for the privilege of life.

To my mother, Iris de Fátima Furtado Pinto, who has always provided me with love and support, and never having tried so hard to achieve my dreams.

To my advisors, Professor Dr. André Ricardo Backes and Professor Dr. Mauricio Cunha Escarpinati, for the friendship, commitment, opportunities, and teachings of this journey.

To Professor Dr. João Fernando Mari for his friendship and contribution to my training.

To Professor MSc. Pedro Moisés de Sousa for all the advice and teachings.

To Professors Dr. Bruno Augusto Nassif Travençolo and Dr. Dalcimar Casanova who kindly accepted to participate in the examination board of this work.

To colleague Jocival Dantas Dias Júnior for the support provided.

To the Federal University of Uberlândia and the Federal University of Viçosa, Institutions that I have provided me with academic and professional growth.

To Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - for financial support. This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil (CAPES) - Finance Code 001. I would like to thank too the company Sensix Inovações em Drones Ltda for providing the images used in the experiments.

*“We are just an advanced breed of monkeys on a minor planet of a very average star.
But we can understand the Universe. That makes us something very special.”
(Stephen Hawking)*

Resumo

Nas últimas décadas, o crescimento populacional e a demanda por alimentos impulsionaram o uso de recursos tecnológicos no que chamamos de Agricultura de Precisão (AP). Dentre essas tecnologias, a utilização de Veículos Aéreos Não Tripulados (VANTs) tem-se destacado por auxiliar, potencializar e otimizar a produção agrícola. As imagens capturadas por VANTs permitem uma visão detalhada da região analisada, visto que o voo ocorre em altitudes baixas e médias (50m a 400m). Além disso, existe uma grande variedade de sensores (câmeras RGB, sensores de captura de calor, câmeras multi e hiperespectrais, entre outros), cada um com características próprias e capaz de produzir informações diferentes. No caso da aquisição de imagens multi espectrais, cada canal da imagem é adquirida por um sensor diferente e em momentos diferentes. Assim, muitas vezes, devido à características intrínsecas ao voo do VANT e a própria natureza da aquisição, dois tipos principais de deformações entre as bandas de uma mesma imagem são comumente observadas, são elas: (1) lineares e (2) não lineares. As deformações lineares consistem, por exemplo, em transformações de rotação, translação e escala; já as não lineares possuem características variadas e sem um padrão previamente definido. Para a correta análise dessas imagens na AP, é necessário identificar e corrigir essas deformações. Diante disso, propõe-se o treinamento de redes neurais profundas para as tarefas de (1) identificação da presença da deformações e (2) a predição dos campos vetoriais de deformação para realizar o registro entre bandas de uma imagem multiespectral. Resultados evidenciam que tais técnicas se mostram promissoras para a solução do problema em questão. Para a identificação da presença de deformação, conseguimos uma precisão de 99,18% em imagens com deformações lineares e 99,25% para deformações não lineares. Além disso, a abordagem proposta alcançou uma precisão de 89,90% a 93,79% na tarefa de estimar o campo de deslocamento entre as bandas. De posse desse campo estimado pela rede, é possível realizar o registro entre as bandas sem a necessidade de marcação manual de pontos.

Palavras-chave: Registro de Imagens, Imagens Multiespectrais, *Deep Learning*, Agricultura de Precisão, VANT.

Detection and correction of deformations in UAVs' images using deep learning

Leandro Henrique Furtado Pinto Silva



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Abstract

In recent decades, population growth and the demand for food have boosted the use of technological resources in what we call Precision Agriculture (PA). Among these technologies, the use of Unmanned Aerial Vehicles (UAVs) has stood out for assisting, enhancing, and optimizing agricultural production. The images captured by UAVs allow a detailed view of the analyzed region since the flight occurs at low and medium altitudes (50m to 400m). In addition, there is a wide variety of sensors (RGB cameras, heat capture sensors, multi and hyperspectral cameras, among others), each with its own characteristics and capable of producing different information. In the case of the acquisition of multi-spectral images, each channel of the image is acquired by a different sensor and at different times. Thus, due to the intrinsic characteristics of the UAV flight and the very nature of the acquisition, two main types of deformations between the bands of the same image are commonly observed, are (1) linear and (2) non-linear. Linear deformations consist, for example, of transformations of rotation, translation, and scale; non-linear ones, however, have varied characteristics and without a previously defined pattern. For the correct analysis of these images in the PA, it is necessary to identify and correct these deformations. Therefore, it is proposed to train deep neural networks for the tasks of (1) identification of the presence of deformations and (2) the prediction of vector deformation fields to perform the registration between bands of a multispectral image. Results show that such techniques are promising for the solution to the problem in question. For the identification of the presence of deformation, we achieved an accuracy of 99.18% in images with linear deformations and 99.25% for non-linear deformations. In addition, the proposed approach reached an accuracy of 89.90% to 93.79% in the task of estimating the displacement field between the bands. With this field estimated by the network, it is possible to register between the bands without the need for manual marking of points.

Keywords: Image Registration, Multispectral Image, Deep Learning, Precision Agriculture, UAV.

List of Figures

Figure 1 – Population growth vs. food production.	12
Figure 2 – Diagram of the light’s electromagnetic spectrum.	14
Figure 3 – Example of an image containing all bands.	14
Figure 4 – Checkerboard of two bands of the same image.	14
Figure 5 – Example of fixed-wings UAV.	18
Figure 6 – Quality factors in remote sensing.	20
Figure 7 – The perceptron: mathematical model of an artificial neuron.	21
Figure 8 – Examples of linearly and non-linearly separable problems.	21
Figure 9 – The MLP structure.	22
Figure 10 – Activation functions.	24
Figure 11 – The structure of a CNN proposed by LeCun et al. (1990).	25
Figure 12 – Lung segmentation using a U-Net.	25
Figure 13 – U-Net architecture.	26
Figure 14 – Necessary steps for registering images using keypoints.	26
Figure 15 – Example of transforming an image to the target image.	28
Figure 16 – The progressive learning architecture proposed by Eppenhof et al. (2019).	30
Figure 17 – Framework proposed by Yasir et al. (2018).	31
Figure 18 – Framework proposed by Dias Junior et al. (2020) and adapted from Yasir et al. (2018).	32
Figure 19 – The flowchart of the proposed co-registration approach by Han et al. (2019).	32
Figure 20 – The construction of both datasets.	36
Figure 21 – Example of patches.	36
Figure 22 – Exemple of cropped region with 64×64 pixels size after a 15-degree rotation transformation.	37
Figure 23 – Perspective transformation in UAV up and down simulations.	37
Figure 24 – Example of subtraction operation between two patches.	38
Figure 25 – Example of non-linear distortion.	39

Figure 26 – Combination of transformations.	39
Figure 27 – Example of non-linear deformation.	40
Figure 28 – Illustration of the architecture of our CNN.	41
Figure 29 – Examples of images from soybean dataset.	46
Figure 30 – Examples of images from cotton dataset.	46
Figure 31 – Example of distortion in near-infrared band.	48
Figure 32 – Combination of two linear transformations.	48
Figure 33 – Example of displacement fields.	49
Figure 34 – The proposed network architecture.	50
Figure 35 – Model accuracy in RGB dataset.	51
Figure 36 – Example of registration in RGB image from the field predicted by our network.	51
Figure 37 – The proposed approach.	52
Figure 38 – Model accuracy in soybean dataset.	53
Figure 39 – Model accuracy in cotton dataset.	54
Figure 40 – Model loss in soybean dataset.	55
Figure 41 – Model loss in cotton dataset.	56
Figure 42 – Example of co-registration in soybean dataset.	56
Figure 43 – Example of co-registration in soybean dataset.	57
Figure 44 – Example of co-registration in soybean dataset.	57
Figure 45 – Example of co-registration in soybean dataset.	58
Figure 46 – Example of co-registration in cotton dataset.	58
Figure 47 – Example of co-registration in cotton dataset.	59
Figure 48 – Example of co-registration in cotton dataset.	59
Figure 49 – Example of co-registration in cotton dataset.	60

List of Tables

Table 1 – Results (%) obtained for our CNN and the compared ones for linear deformations.	42
Table 2 – Results (%) obtained for each CNN model for non-linear deformations. .	43
Table 3 – Number of parameters of each CNN model.	43
Table 4 – Misalignment average, in pixels, between the sensors present in soybean dataset.	47
Table 5 – Misalignment average, in pixels, between the sensors present in cotton dataset.	47
Table 6 – Parameters for B-spline transformations.	47
Table 7 – Summary of validation accuracy for each of the moved bands considered.	53

Acronyms list

BRISK Binary Robust Invariant Scalable Key-points

BPE Back Projection Error

CP Conjugate Points

CNN Convolutional Neural Network

CT Computed Tomography

DLIR Deep Learning for Image Registration

DS1 Data Set 1

DS2 Data Set 2

GNSS Global Navigation Satellite System

GSD Ground Sample Distance

IBGE The Brazilian Institute of Geography and Statistic

MLP Multi-Layer Perceptron

NIR Near-Infrared

ORB Oriented fast and Rotated Brief

PA Precision Agriculture

RGB Red, Gren, and Blue

ReLU REctified Linear Units

SIFT Scale Invariant Feature Transform

SURF Speeded-Up Robust Feature

UAV Unmanned Aerial Vehicles

UN United Nations

Contents

1	INTRODUCTION	11
1.1	Problem Definition	13
1.2	Goals	15
1.3	Hypothesis	15
1.4	Contributions	15
1.5	Thesis Organization	16
2	FUNDAMENTALS	17
2.1	Precision Agriculture	17
2.1.1	Unmanned Aerial Vehicles in Precision Agriculture	18
2.1.2	Remote Sensing in Precision Agriculture	18
2.2	Neural Networks	19
2.2.1	Activation functions	22
2.3	Convolutional Neural Networks	23
2.3.1	U-Net	24
2.4	Image Registration	25
2.4.1	Feature Detection	27
2.4.2	Feature Matching	27
2.4.3	Construction of Mapping Function	27
2.4.4	Image Transformation	28
3	RELATED WORKS	29
4	DEFORMATION RECOGNITION IN UAVS' IMAGES USING DEEP LEARNING	35
4.1	Image Datasets	35
4.1.1	Linear deformation dataset	36
4.1.2	Non-linear deformation dataset	38

4.2	Experiments	40
4.3	Results and Discussions	41
5	DEFORMABLE CO-REGISTRATION USING DEEP LEARN- ING	45
5.1	Image Dataset	45
5.2	Data preparation	45
5.3	End-to-end training	49
5.4	Experimental Results in RGB Dataset	50
5.5	Experimental Results in Multispectral Datasets	52
5.6	Discussions	54
6	CONCLUSION	61
6.1	Main contributions	61
6.2	Limitations	62
6.3	Future Works	62
6.4	Contributions in Literature	62
	BIBLIOGRAPHY	65

CHAPTER **1**

Introduction

Humanity, since the beginning, has gone through a process of population growth, and it has been accelerated over the last century. Proof of this is that it is estimated that the number of inhabitants of the planet two thousand years ago was close to 250 million. In 1650, the number grew to 500 million. The 1 billion people mark was only reached in 1850. The mark had already been doubled a century later, when, in 1950, the population was approximately 2.5 billion people (PENA, 2013). Currently, the United Nations (UN) estimates that the global population is about 7.8 billion people (UN, 2020b).

The work of Thomas Robert Malthus, a British economist, entitled “Essay on the Principle of Population” (1798), which coined a very widespread theory at the time, called the Malthusian Theory, was quite pessimistic. Malthus believed that poverty would be part of the destiny of humanity, based on the assumption that the human population had unlimited growth potential, unlike food production, which had limits. Malthus claimed that while the population grew according to a geometric progression, food production grew according to an arithmetic progression (MALTHUS; WINCH; JAMES, 1992). At the end of the 19th century, several studies also indicated a concern with the growth of the world population and the planet’s capacity to produce food. These studies pointed to a future scenario of drastic food shortages and, as a consequence, hunger. Figure 1 presents a summary of the relationship between population growth and food production capacity.

The forecasts were not confirmed, largely due to the significant technological advances that occurred in the agricultural area in the 1950s and 1960s. A set of technology development initiatives that made it possible to increase agricultural production worldwide, especially in developing countries. The initiatives resulted in the adoption of new technologies, including varieties of high-yielding cereals, mainly wheat and rice, the use of chemical and agrochemical fertilizers, irrigation methods, new forms of cultivation, and mechanization of crops (MALTHUS, 1872; HAZELL, 2009; FARMER, 1986).

In Brazil, the agribusiness sector is increasingly consolidated. Proof of this is that amid the SARS-CoV-2 pandemic (“coronavirus”), declared by the UN on March 11, 2020, the

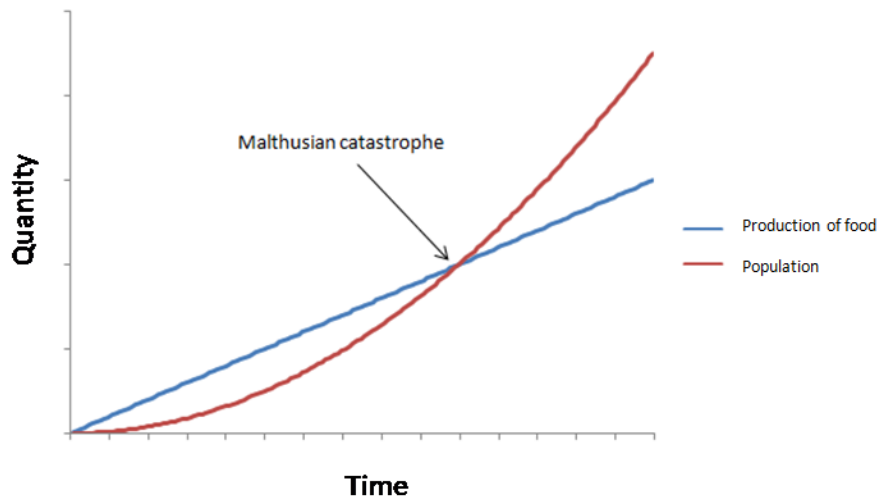


Figure 1 – Population growth vs. food production: projections pointed to population growth in geometric progression, while food production capacity was growing in arithmetic progression.

agribusiness sector was the only one to show growth, according to the The Brazilian Institute of Geography and Statistic (IBGE). The growth, although timid (about 0.6 %), in the first quarter of 2020, shows the strength of the sector, even in a traumatic period for the national and international economy (UN, 2020a; IBGE, 2020a; IBGE, 2020b).

One of the main components of the current phase of agricultural research is Precision Agriculture (PA), which can be characterized by agricultural management based on observation, measurement and response to inter and intra-field variability in crops. The objective of the research in PA is to define a Decision Support System for the management of the entire farm, to optimize the return of inputs and preserving resources. The area of PA proved to be strongly dependent on imaging and mapping technologies to estimate growth or identify other important agronomic characteristics, such as nitrogen stress. Given this scenario, with the increasing importance of using images in the PA, some mechanisms for obtaining these images have stood out, in particular, Unmanned Aerial Vehicles (UAV).

Advances in UAV technology, especially over the last decade, have led to its wide dissemination in the most diverse segments, including precision agriculture. With the consequent reduction in production costs and droppin operating costs, small and medium producers can now pay for the use of image-assisted technologies, such as UAV (JENKINS; VASIGH, 2013).

Unlike other aerial imaging devices, such as satellites and large aircraft, UAV capture images at low and medium altitudes (50 to 400 m), providing a more detailed view of the analyzed region. Another important element for the efficiency of the analyzes carried out with this equipment are the sensors used. There is a wide variety of devices used in the

process: Red, Green, and Blue (RGB) cameras, heat capture sensors, multi and hyperspectral cameras, among others. Each device has its characteristics and produces information that requires different types of analysis. However, the data acquisition process, in general, is the same regardless of the sensor used: the equipment is attached to the aircraft and images are captured sequentially during the flight. After the end of the process, these images are organized in a mosaic representing the entire area (MCBRATNEY et al., 2005; MILELLA; REINA; NIELSEN, 2019; BLACKMER; SCHEPERS, 1996; SANKARAN et al., 2015; KATAOKA et al., 2003).

A multi or hyperspectral image, to be useful in agricultural applications, needs, first, to have its bands aligned. This process, which is known as co-registration, presents a series of difficulties to be performed (JUNIOR; BACKES; ESCARPINATI, 2019). The first difficulty to consider is the fact that each element present in the image is represented differently in each spectrum, due to chemical and physical characteristics. Emission, absorption, reflection, and transmission processes vary for each spectrum. Therefore, there is a loss of information and, consequently, difficulty in the alignment process (BANERJEE; RAVAL; CULLEN, 2018).

1.1 Problem Definition

Although the process of acquiring multispectral images, especially at low/medium-altitude, has become simpler, the challenge of aligning the different bands, especially when there are non-linear deformations, is a major challenge.

The mentioned situation is due to the simple fact that current techniques are consolidated and robust in more stable image processing problems and with greater spatial coverage when compared to those obtained by UAV (BANERJEE; RAVAL; CULLEN, 2018; BERNI et al., 2009; HUNT et al., 2010). As mentioned, the objects of study of this work are multi-spectral images. In other words, each band covers a different range of the electromagnetic spectrum (see Figure 2). In this way, objects that are represented by one spectrum may be not represented in others. This situation impairs the task of finding control points and makes the task more complex. Also, the complexity of the task still increases considerably when we have distortions between these bands, especially non-linear ones (JUNIOR et al., 2020; JUNIOR; BACKES; ESCARPINATI, 2019).

Figure 3 shows the five separate bands of the same image captured by UAV in a soybean plantation. Subsequently, Figure 4 presents an image composed of two spectra, through a checkerboard (where light and dark tones differentiate different spectra). In this figure, it is possible to see the misalignment between the spectra of the same image. In other words, the simple overlapping of the bands would represent a completely misaligned multispectral image.

Also, the use of UAVs to obtain agricultural images requires hundreds or, in many

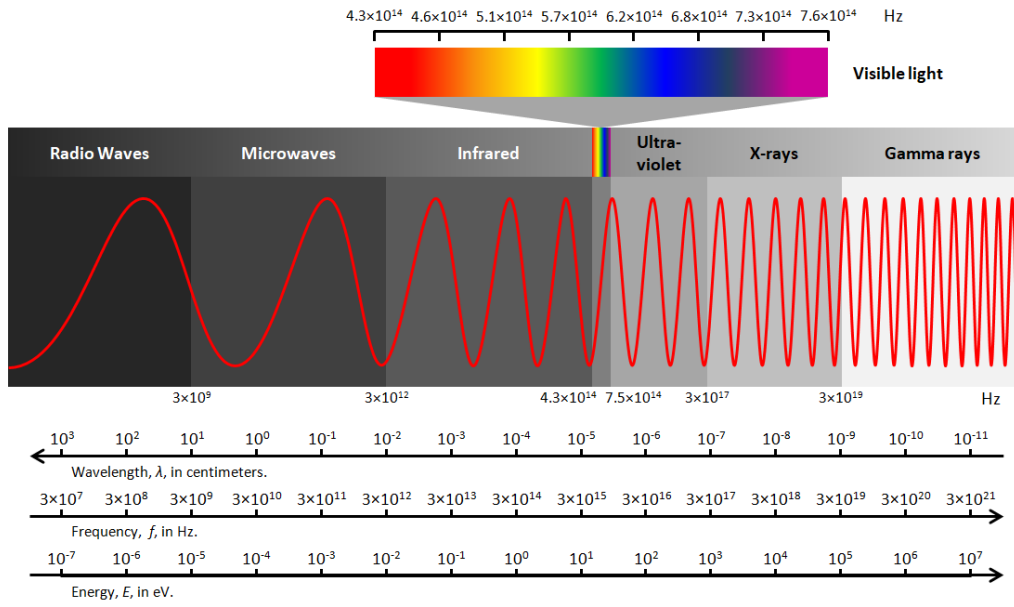


Figure 2 – Diagram of the light’s electromagnetic spectrum, showing the different wavelengths.

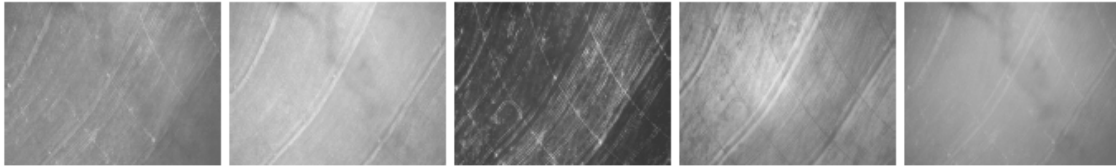


Figure 3 – Example of an image containing all bands. Blue, Green, Red, Near-Infrared, and Red Edge, respectively. Adapted from (JUNIOR et al., 2020).

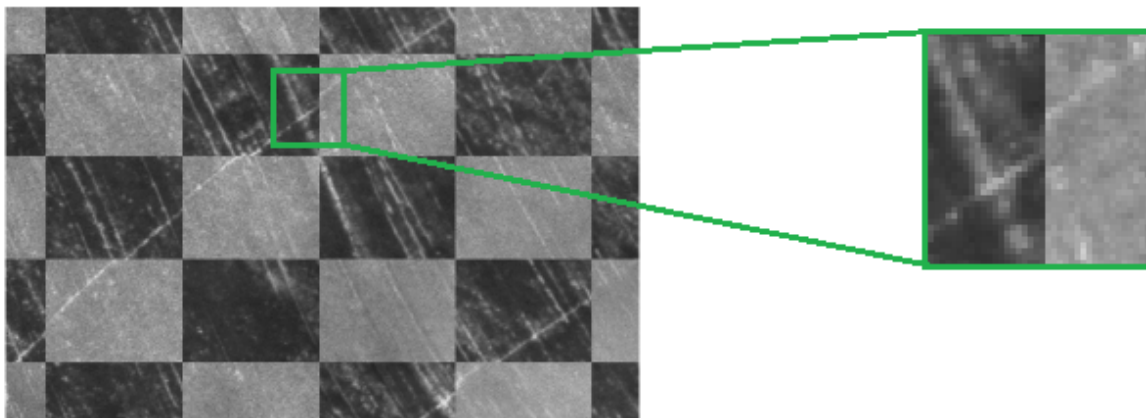


Figure 4 – Checkerboard of two bands of the same image. Note that in highlight there is a misalignment between the bands, where the same line, in its different bands, should follow the same diffraction (JUNIOR et al., 2020).

cases, thousands of overlaps to cover a given area, which demands considerably robust

techniques to align the bands of each of these images.

1.2 Goals

The main goal of this work is to investigate the ability of Convolutional Neural Network (CNN) to identify and correct linear and non-linear deformations in images obtained by UAVs in PA.

More specifically:

- ❑ Identify the main deformations in images obtained by UAVs, which are inherent to the nature of its flight.
- ❑ Propose an approach based on deep learning that can recognize the presence of these deformations.
- ❑ Compare the proposed approach with traditional CNN for the deformation identification problem.
- ❑ Propose an approach based on deep learning to correct the deformation and carry out co-registration.
- ❑ Evaluate the network's ability to perform co-registration.

1.3 Hypothesis

This work is based on two main hypotheses:

1. Deep learning is capable of detecting the presence (or not) of different types of deformations inherent to the UAV flight, whether these are linear or non-linear deformations.
2. From the identification of deformations, a deep learning approach is also able to estimate the displacement field between a fixed image (or band) and one considered moved. With this displacement field predicted by the network, it is possible to apply it on the moving image towards the fixed image and perform the registration without having to manually marked points.

1.4 Contributions

The main contributions of this work are:

- ❑ The assessment regarding the ability of deep learning to identify deformations in images obtained by UAVs.

- ❑ A framework based on deep learning for performing co-registration considering linear and non-linear deformations, without the need (after training) of manual scoring of points to perform the co-registration task.

1.5 Thesis Organization

This thesis is organized as follows:

- ❑ In Chapter 2 we present the theoretical basis of this work, with the presentation of concepts related to Precision Agriculture, Unmanned Aerial Vehicles, Remote Sensing, Neural Networks, and Image Registration.
- ❑ In Chapter 3 we present the state-of-the-art related to our problem, which are the basis for the construction of the present work.
- ❑ In Chapter 4 we present an approach based on deep learning to identify the presence (or not) of linear and non-linear deformations in aerial images obtained by UAVs.
- ❑ In Chapter 5 we propose an deep learning based approach to estimate the displacement field between the bands of a multispectral image obtained by UAV to perform the co-registration.
- ❑ Finally, in Chapter 6 we present the conclusions of this work, limitations, future works, and contributions to the literature from this work.

Fundamentals

In this chapter, concepts, techniques, procedures, and approaches that underlie the present work will be presented. In particular, we will exhibit on Precision Agriculture, Unmanned Aerial Vehicles, Remote Sensing, Neural Networks, and Image Registration.

2.1 Precision Agriculture

The demand for food will increase in the coming decades. Estimates show that the world population will be close to 10 billion people by 2050. However, through increasingly advanced techniques in PA, it is expected to double food production for the same period mentioned, using only 5% more of the currently planted area. In this way, agricultural production would be made smartly, optimizing natural resources and avoiding drastic environmental changes (HUNTER et al., 2017).

Traditional agriculture treats the fields homogeneously, while the PA aims to divide the fields into management areas that are treated individually. Thus, each zone receives its inputs according to their needs and characteristics. This results in the optimization of resources, generating profit for farmers since the inputs are applied precisely, avoiding waste (MAES; STEPPE, 2019).

A great example of success in PA is the Netherlands. The small European country, whose territory does not exceed 42,000 km², managed in 2017 to be the 2nd largest exporter of agricultural products in the world. Its exports totaled 112 billions dollars, which placed the country behind only the United States and ahead of Germany (3rd place) and Brazil (4th place). Due to the small territory, the Netherlands cannot carry out large-scale agriculture, like the USA and Brazil, and so the country specializes in high-performance agriculture, with average yields three times higher than its European neighbors (DALL'AGNOL; MOREIRA, 2019).

In short, PA is considered to be the third revolution in modern agriculture. The first revolution (1900 to 1930) was the mechanization of farms. The second revolution (1990 to 2005) was the genetic modification of plants (JUNIOR et al., 2020).

2.1.1 Unmanned Aerial Vehicles in Precision Agriculture

In the early 2000s, satellites provided images with low spatial and temporal resolutions. Thus, the images generated incorrect analyzes and consequent waste of agricultural inputs. Subsequently, airplanes and balloons were used to obtain agricultural images. These images provided improvements in the resolutions, however, the cost to produce this kind of images was higher. Currently, UAV have become popular due to their more affordable cost and the provision of high-resolution images (MILICS, 2019). These factors have led to UAV being widely used by farmers, as well as generating interest in academic research. According to Maes e Steppe (2019), the number of papers that used UAV in PA has grown exponentially in the last 10 years.

UAV are divided into two main classes, each with its advantages: (1) multi-rotor, and (2) fixed-wing. The advantages of multi-rotor over fixed-wings consist of greater maneuverability, more compactness, ease of use, and lower price. The advantages of fixed-wings are their range since this type of UAV performs a linear flight with greater stability. Figure 5 shows example of fixed-wings UAV (JUNIOR et al., 2020).



Figure 5 – Example of fixed-wings UAV (JUNIOR et al., 2020).

One of the main characteristics of UAV is the possibility of coupling several sensors and thus making specific measurements. With these measures in place, we can assess, for example, the level of soil nutrients, disease detection, and dry stress. An overview of the possibilities for evaluating the farm through remote sensing, especially through UAV.

2.1.2 Remote Sensing in Precision Agriculture

Remote sensing can be defined as obtaining information about a particular object without physical contact. In PA, specifically, it consists of a field-level data capture

process. Initially, the images were obtained by satellites, airplanes or helicopters (ZARCO-TEJADA et al., 2016). However, these processes did not provide data quickly enough for the “ideal” use of PA methods. Also, they were considerably expensive.

On the other hand, especially in the last decade, the development and popularization of UAV has fostered a new era of PA. These devices can provide unprecedented spatial, spectral and temporal resolution data (MAES; STEPPE, 2019; DYRMANN; KARSTOFT; MIDTIBY, 2016).

In the PA, remote sensing images have the most varied uses: classification of crop species, monitoring of diseases and weeds, detection of dry stress and mapping of soil properties (IHUOMA; MADRAMOOTOO, 2017). However, some aspects need to be evaluated before using remote sensing images for decision-making processes. Several parameters are affected by the type of platform (satellite, aerial or terrestrial) and the sensor chosen for data acquisition: (1) geometric accuracy of the images; (2) radiometric; (3) spectral; (4) spatial; (5) temporal resolution; (6) width and the number of spectral bands; (7) quality of the spectral information represented in the images (KHANAL; FULTON; SHEARER, 2017).

2.2 Neural Networks

Neural networks are computer systems designed based on the human brain study. Thus, neural networks are different from other existing computational models. Such differentiation is because the neural paradigm does not make use of the concepts that until then characterized other computer systems algorithms (FAUSETT, 2006).

The main characteristic of neural networks is the ability to learn based on the presentation of examples. Thus, the construction of a neural network is based on the configuration of its internal architecture (an interconnected network of neurons) and on the training of that network, based on examples, so that it can learn how to solve problems (FAUSETT, 2006).

Neural Networks use a mathematical model to simulating the basic functioning of the brain. In this sense, the intensity of the connections between neurons is stimulated through weights, which are adjustable during the network training. The neurons are activated through the composition of two functions: activation and propagation. The purpose of these functions is to map the input signals to a single output signal. Such an output signal will then be propagated to the next neurons in the Neural Network (FAUSETT, 2006).

There are several models of neural networks in the literature. Each of these models comes from different lines of research and aims at better performance in solving a specific problem (FAUSETT, 2006). These neural models can be based on:

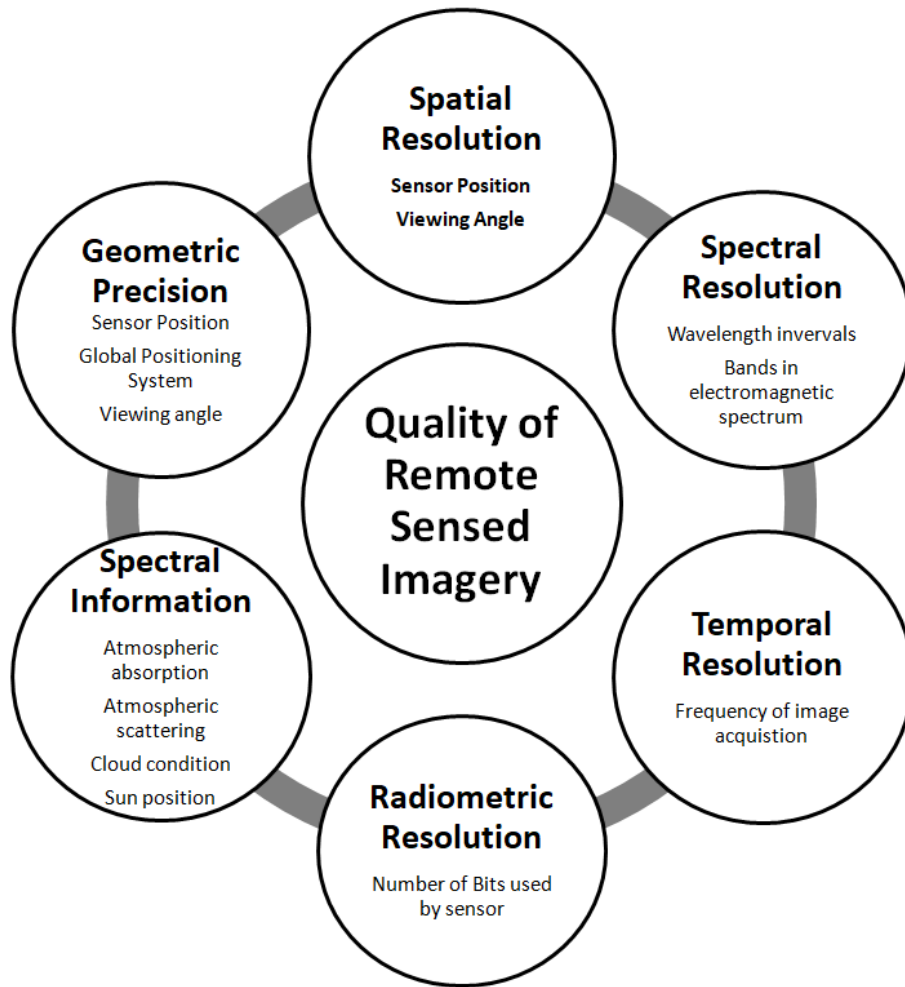


Figure 6 – Quality factors in remote sensing. Adapted from (KHANAL; FULTON; SHEARER, 2017).

- ❑ **The training strategy:** (1) supervised, when there is an explicit target and the network aims to reach it; or (2) unsupervised, when the network seeks to identify clusters or patterns from the data, without a predefined target.
- ❑ **The form of training:** (1) incremental, which occurs when weights are adjusted after the presentation of each input pattern; or (2) in batch, when the weights are adjustment will only be carried out after the stimuli.
- ❑ **The form of operation:** (1) unidirectional, where the internal signals will propagate only in the direction of input/output; or (2) recurring, when there is feedback from this process.

An artificial neuron is the fundamental unit for processing information of an artificial neural network. The mathematical model of an artificial neuron was conceived by researchers W.S. McCulloch and W.H. Pitts, in 1943. This model, also known as Perceptron, is composed of connections simulating dendrites, weights simulating synapses, a

mapping function simulating the cell body, in addition to an output that simulates the axon (FAUSETT, 2006).

Figure 7 illustrates the Perceptron, where x are the inputs and w are the weights. There is a sum function relating the entries with their respective weights to be inserted in an activation function. Depending on the threshold reached by this function, the neuron is (or not) activated for an output y .

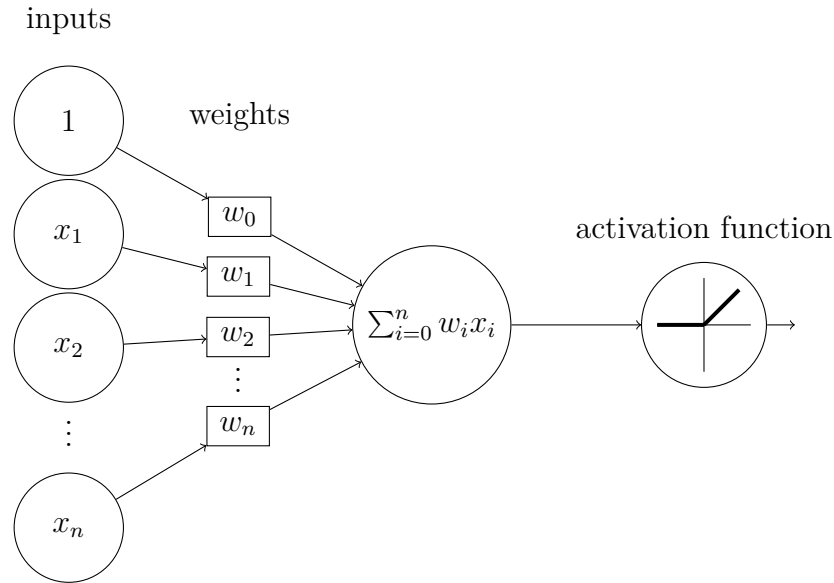


Figure 7 – The perceptron: mathematical model of an artificial neuron.

Perceptron enables us to learn binary classifications, i.e., those that are linearly separable. A problem is linearly separable when we can define two distinct classes on the two-dimensional Cartesian axis by drawing a line (if we consider more dimensions, this separation is through a hyper-plane). When we are unable to separate classes using a straight line, we have a non-linearly separable problem. Figure 8 shows examples of linearly and non-linearly separable problems.

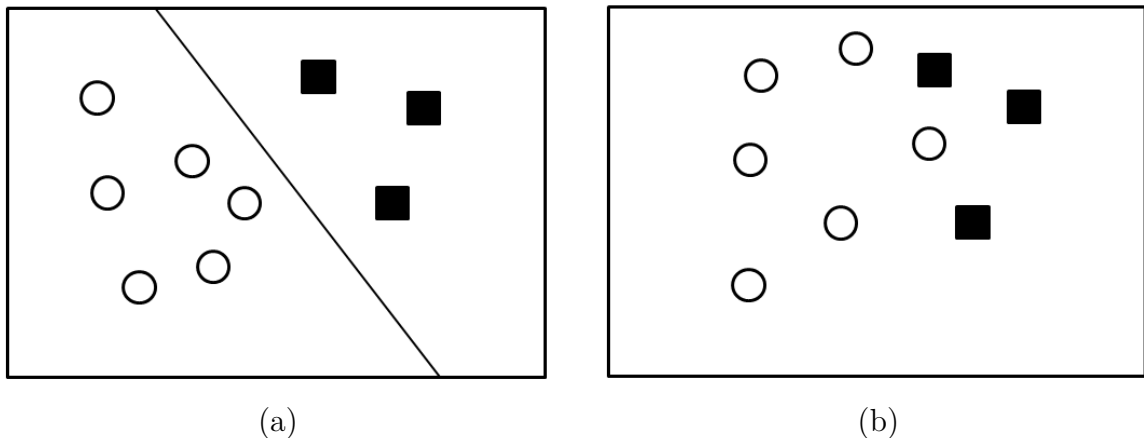


Figure 8 – Examples of (a) linearly and (b) non-linearly separable problems.

To address non-linearly separable problems, we can make use of a Multi-Layer Perceptron (MLP), a network composed of multiple layers of perceptrons. In a MLP, learning is usually done through the error backpropagation algorithm. In short, to train a neural network through error backpropagation, it involve three stages: (1) the training pattern feedforward, (2) the calculation and the backpropagation of the associated error, and, finally, (3) adjusting the weights. In other words, the coefficients of the weights used to propagate the errors backward are the same as those used when calculating the output. Only the direction of flow is changed, and this technique is applied to all layers of the network. According to Fausett (2006), after the training, the network solution is obtained through the calculation of the feedforward phase.

Succinctly, MLP is organized in a structure of fully-connected layers, in the feedforward sense. A neuron located in a given layer has its output connected to all neurons in the next layer and no connections to the previous layers (HASSOUN et al., 1995). Figure 9 illustrates this topological structure.

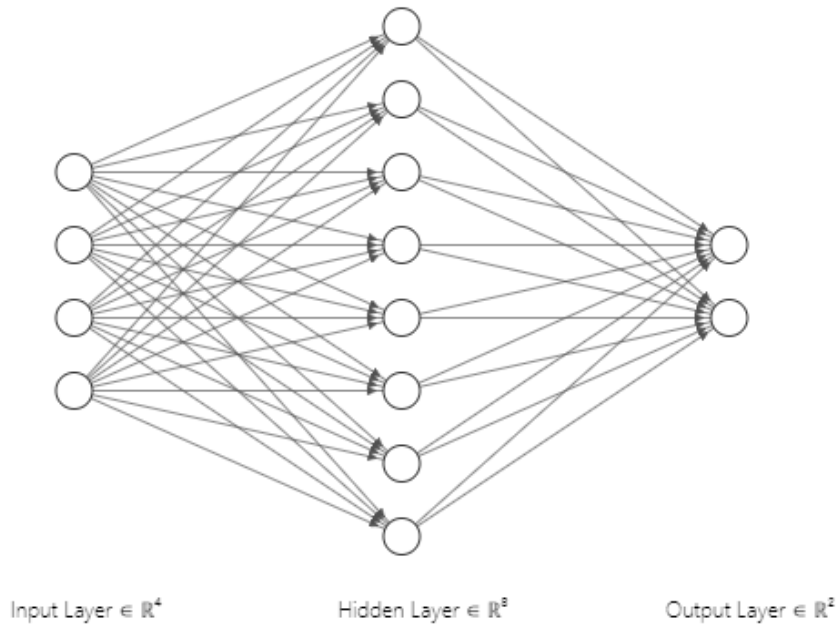


Figure 9 – The MLP structure.

2.2.1 Activation functions

An activation function is an algebraic operation responsible for defining whether or not a given neuron should be activated, through an algebraic operation (HAYKIN; NETWORK, 2004; GOODFELLOW et al., 2016). Among the main activation functions, we highlight Linear, Sigmoid, Hyperbolic Tangent, and REctified Linear Units (ReLU), which will be exposed below and are illustrated by Figure 10.

- **Linear:** simple activation function, which consists of applying the identity operation to the data. In this function, the output data is proportional to the input data. In general, neuron activation consists of the weighted sum of the input of this neuron. They are commonly used for neurons in the input layer in a neural network. This function is defined as: $f(x) = x$.
- **Sigmoid:** used for generating outputs whose values vary between 0 and 1. It is sensitive to saturation when the argument is in positive and negative extremes. Thus, this function is not sensitive to small changes, which makes training more difficult. This function is defined as: $\sigma(x) = 1/(1 + e^x)$.
- **Hyperbolic Tangent:** a function that resembles the identity function, being an alternative to the Sigmoid function for activating hidden layers of neural networks. This function is defined as: $TanH(x) = 2\sigma(2x) - 1$.
- **ReLU:** the most popular function for feed-forward neural networks because of its simplicity to enhance faster training. In short, the output is 0 if the input is less than or equal to 0; and the output is equal to the input if the input is greater than 0. This function is defined as: $ReLU(x) = \max(0, x)$.

2.3 Convolutional Neural Networks

CNN are a category of deep learning algorithms inspired by the human learning process. These networks are based on the concept of receptive field of biological systems, which gives these networks the ability to learn different filters and characteristics of an image. In this way, CNN can explore the spatial correlations between pixels in an image to extract relevant image attributes for different tasks, such as image classification and segmentation (LECUN et al., 1998; GUO et al., 2016; PONTI et al., 2017).

Most CNN models available in the literature are defined in terms of three types of layers (convolutional, pooling, and dense), which are combined differently to improve image classification or segmentation.

The convolutional layer is responsible for extracting significant attributes from an image. For this, a series of convolution operations are applied to the input data, which act as receptive filters and, in this way, different attributes of a local region of the image stand out. A ReLU activation function and, generally, a Batch Normalization operation, depending on the architecture in question, are usually applied to the result of the convolutional layer. This helps to speed up network training and improve the results (LECUN; BENGIO; HINTON, 2015).

Convolutional layers are generally followed by a pooling layer, whose objective is to reduce the characteristic maps calculated by the previous layers, thus reducing the net-

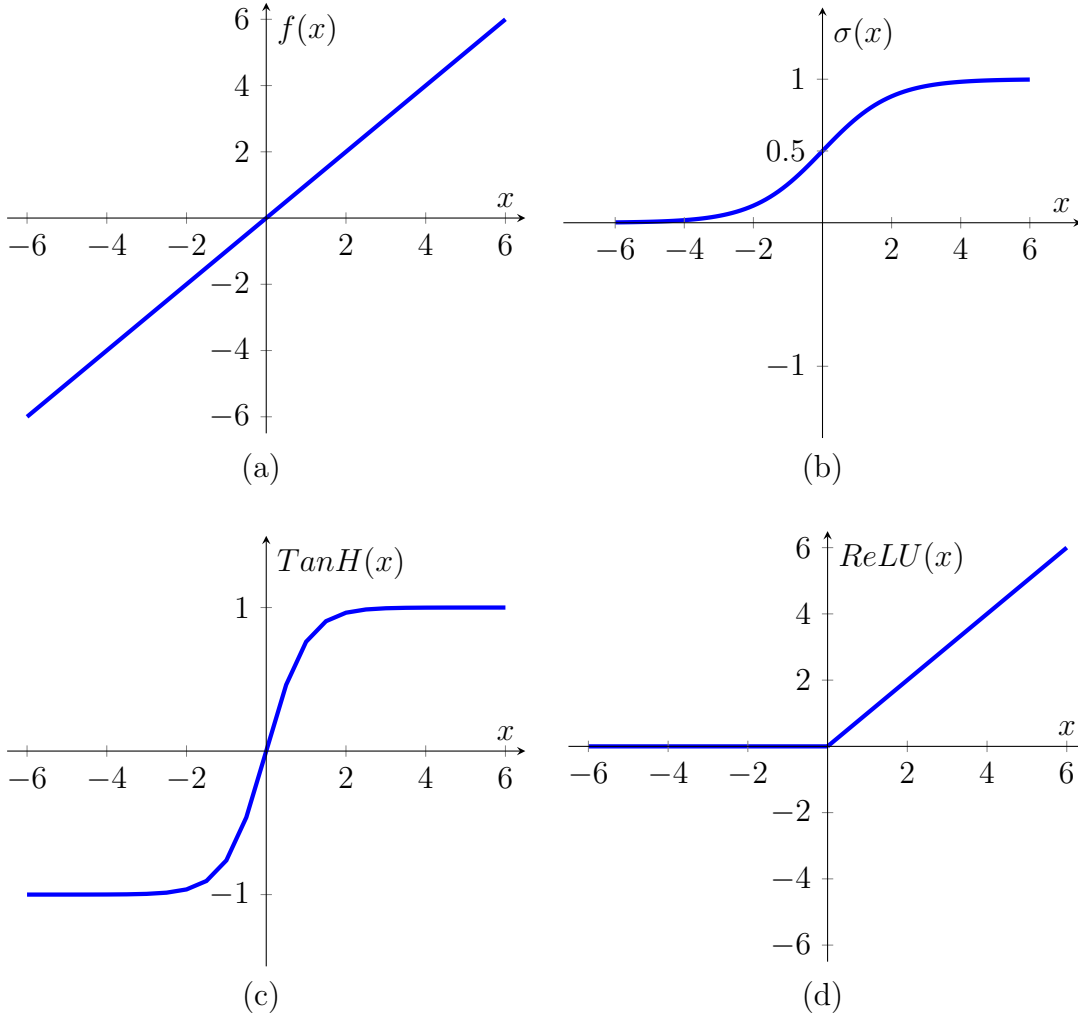


Figure 10 – Activation functions: (a) Linear, (b) Sigmoid, (c) Hyperbolic Tangent, and (d) ReLU.

work's sensitivity to image distortions and data changes. In general, a mask of size 2×2 is used, thus reducing a region of 4 pixels to a single value according to some criterion (for example, maximum or average pixel of the region).

At the end of CNN, we find the dense layers. These layers receive the 2D feature maps obtained from the previous layers as input, to learn a vector of 1D features capable of discriminating the input image. This vector is used as input for a softmax classifier, which returns the most likely class for a given input image.

Figure 11 shows a classic CNN structure, which was proposed by LeCun et al. (1990). It is possible to observe the convolution, pooling, and dense layers.

2.3.1 U-Net

U-Net is an architecture inspired by the Fully Convolutional Network, proposed to address segmentation problems in medical images (see Figure 12) (LONG; SHELHAMER; DARRELL, 2015). The main idea of U-Net is to replace pooling operations with up-

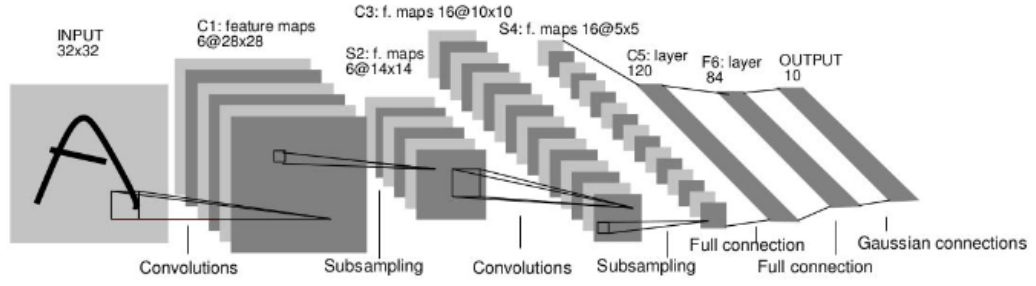


Figure 11 – The structure of a CNN proposed by LeCun et al. (1990).

sampling operations. In this way, these layers can increase the output resolution. Also, the use of a successive convolutional layer may be able to learn to assemble an accurate output, based on the input information (RONNEBERGER; FISCHER; BROX, 2015).

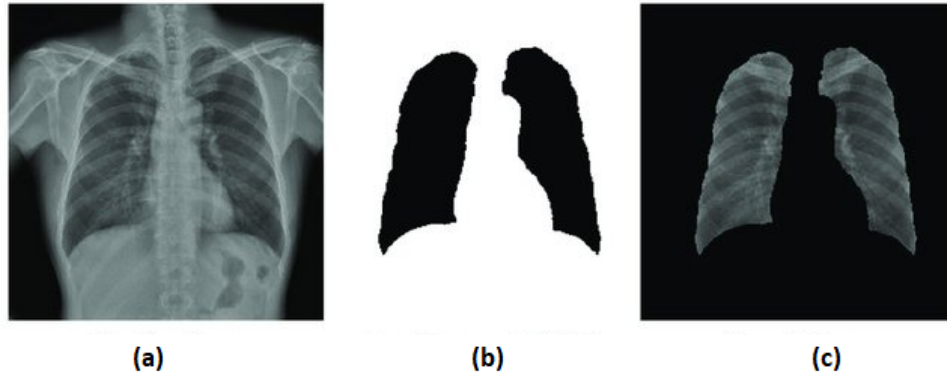


Figure 12 – Lung segmentation using a U-Net: in (a) we have a chest X-ray image; in (b) we have a area to be segmented by U-Net; in (c) we have a segmented image (HEO et al., 2019).

The network consists of an encoding and decoding path analogous to the letter *U* (see Figure 13). In the encoding path, we have a typical convolutional network, where there is the successive application of convolutions, with the ReLU activation function, followed by a pooling operation. The decoding path combines features and spatial information, through a sequence of convolutions and concatenations with features, originating from the high-resolution coding path for reconstructing the segmentation map (BAHETI et al., 2020; IGLOVIKOV; SHVETS, 2018).

2.4 Image Registration

The image registration process consists of a geometric transformation, which relates the coordinates of a reference image *A* to the coordinates of an image to be registered *B*. This transformation aims to eliminate existing distortions in image *B*, aligning it towards *A*.

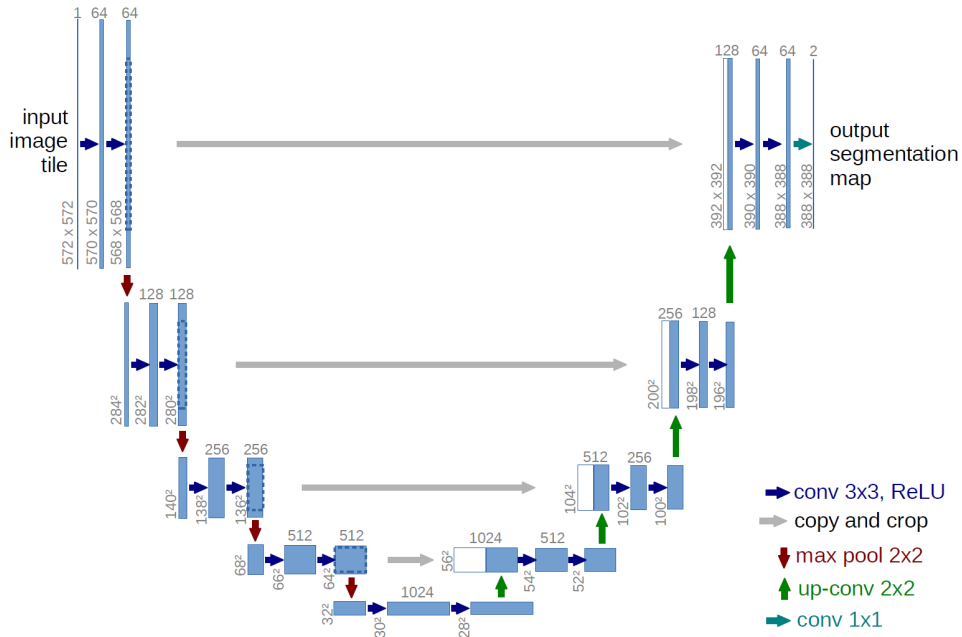


Figure 13 – U-Net architecture (RONNEBERGER; FISCHER; BROX, 2015).

The vast majority of methods for image registration are composed of four basic steps: (1) Feature Detection, (2) Feature Matching, (3) Construction of Mapping Function, and (4) Image Transformation (UCHIDA, 2013). Figure 14 illustrates these four steps.

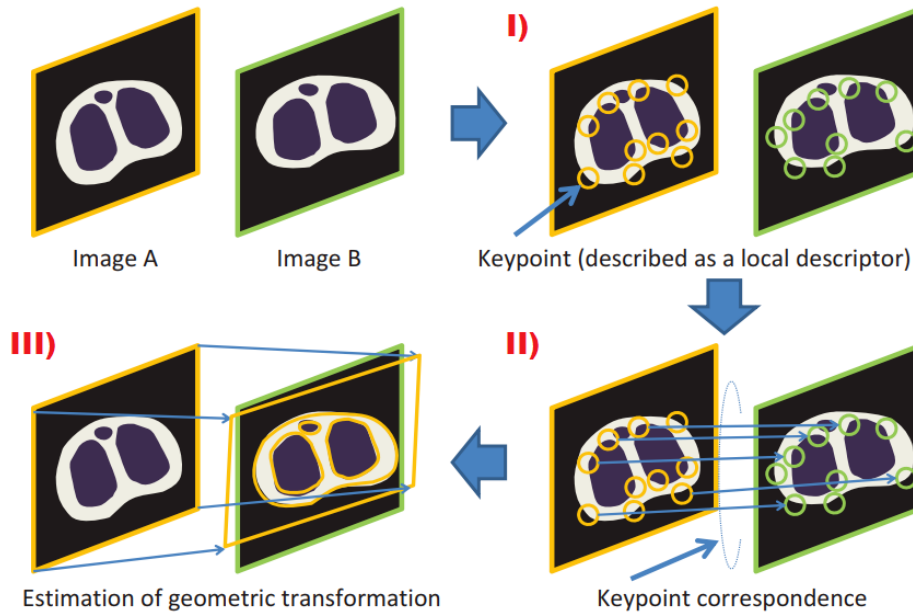


Figure 14 – Necessary steps for registering images using keypoints. In (I) the extraction of features from the image; in (II) the combination of features; and in (III) the mapping function and the transformation of that image (UCHIDA, 2013).

2.4.1 Feature Detection

Features can be understood as a pattern that takes place in a specific location in the image and they are distinguished from its closest neighbors. In general, such a pattern is associated with a sudden change in one (or several) properties of an image (texture, color, or intensity, for example). The registration algorithms that use keypoints find such points by analyzing the magnitude and direction of the intensity changes in the local neighbors of the image, in order to detect regions, corners, and edges (ANDERSSON; MARQUEZ, 2016). Traditional algorithms for detecting features are Scale Invariant Feature Transform (SIFT), Speeded-Up Robust Feature (SURF), Binary Robust Invariant Scalable Keypoints (BRISK), and Kaze Features (LOWE, 2004; BAY; TUYTELAARS; GOOL, 2006; LEUTENEGGER; CHLI; SIEGWART, 2011; ALCANTARILLA; BARTOLI; DAVISON, 2012).

2.4.2 Feature Matching

After extracting the keypoints, it is necessary to perform the correspondence between the points obtained in each of the images. Two approaches can be applied to obtain this correspondence: exhaustive and approximate methods (BANERJEE; RAVAL; CULLEN, 2018).

In an exhaustive method (brute force, for example), for each keypoint obtained in the first image a distance measure (Hamming for binary images and Euclidean for the others (BANERJEE; RAVAL; CULLEN, 2018)) is calculated for all points in the second image. This method seeks to find the closest correspondence for each analyzed point. This approach has a high computational cost due to the excessive amount of comparisons made.

The approximate methods provide a large computational gain with only a small loss of accuracy. In this sense, depending on the application and necessary accuracy, the approximate methods can be used to obtain the correspondence between the points. Among the algorithms used for this purpose, we can highlight Hierarchical, K-Means Tree, and Automatic Selection of the Optimal (BANERJEE; RAVAL; CULLEN, 2018).

2.4.3 Construction of Mapping Function

A mapping function can be mathematically defined as a 2D function which maps the (x, y) coordinates of a given image A to the (x, y) coordinates of an image B . Two main types of function of mapping, Linear and Non-Linear, are characterized by the type of deformation in the image (GONZALEZ; WOODS, 2006).

2.4.4 Image Transformation

The last step to carry out the registration process is to transform the image according to the transformation function obtained. In this sense, to transform the images according to a target image, other images are mapped according to that transformation.

This transformation is carried out by translating the points marked on the moving image towards the corresponding points on the fixed image. The mentioned process is illustrated in Figure 15.

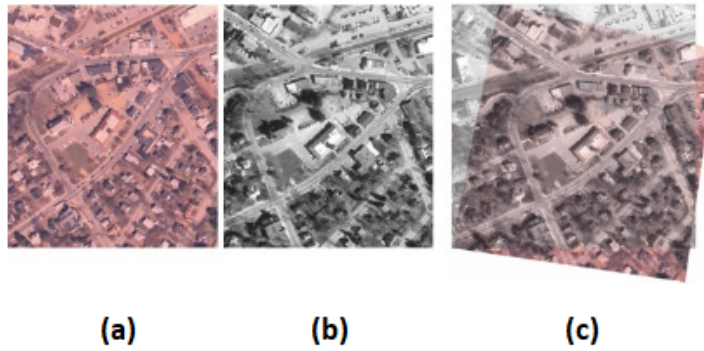


Figure 15 – Example of transforming an image (a) to the target image (b). The result of the overlap, after transformation, is shown in (c) (MATHWORKS, 2018).

Related Works

In this chapter, we present the works that relate to this one, which addresses computational techniques and methods applied to images obtained by UAV in PA. These works presented in this chapter form the basis for what is proposed by this work.

In the paper of Eppenhof et al. (2019) it was proposed to use deep learning methods for registration images with non-linear distortions as an alternative to traditional registration methods. For the complex scenario in question, a multiple resolution task is required. Thus, instead of training a large CNN in the task of a single registration, smaller versions of the network were initially trained with low-resolution images and less complex deformation fields. During training, the network was progressively expanded with additional layers, which were trained with high-resolution data.

Figure 16 shows the progressive training proposed by Eppenhof et al. (2019). The architecture in question is an adaptation of U-Net, and the input is an image with two channels, one of which has the fixed image and the other with the displaced image. The output consists of the vector fields of displacements for each axis of the image. In their work the images have three dimensions, so the output will be the respective deformation fields in x , y , and z .

The results showed that this training approach allows a network to learn larger displacement fields without sacrificing registration accuracy and that the resulting network is less prone to registration errors compared to running the entire network at once. The authors also agreed that a progressive training procedure leads to greater accuracy of the registration when learning large and complex non-linear deformations.

The paper of Yasir et al. (2018) proposes the creation of a computational framework for multispectral image registration. This framework is based on having a minimum amount of correspondence between control points between two channels.

To obtain the key-points, the SIFT, BRISK, SURF, and Oriented fast and Rotated Brief (ORB) techniques were used. This framework was evaluated using the number of control points obtained, the computational time, and the Back Projection Error (BPE). In addition, the robustness of the framework in three databases was evaluated, two with

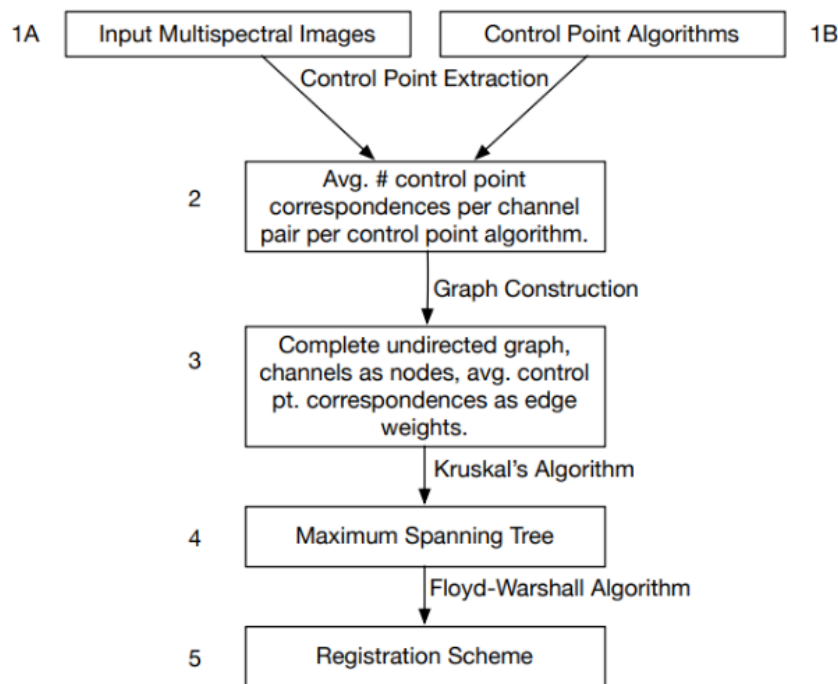


Figure 17 – Framework proposed by Yasir et al. (2018).

and with other methods in the literature. The proposed framework has a BPE of 0.425 pixels, which is 335% higher than the evaluated frameworks.

Figure 18 presents the framework proposed by Dias Junior et al. (2020). It is possible to note the adaptation of Yasir et al. (2018) initial proposal, specifically for agricultural images. The same figure also shows the indication of the “green” channel as something, that is, a fixed channel to which the others must be moved in their direction, to correct misalignments.

The paper of Han et al. (2019) addresses the problem of co-registration in very high-resolution images obtained through UAVs. The authors present an automated co-registration approach, which is illustrated in Figure 19.

In short, the approach makes use of the spatial information obtained at the time of capture of the images provided by the Global Navigation Satellite System (GNSS), in order to assist in the correlation-based coregistration phase. Then, a transformation function between multispectral images is estimated based on Conjugate Points (CP), which are obtained locally on the images using the phase correlation approach.

The validation of the method proposed by Han et al. (2019) considered two datasets containing multispectral images. The proposed method was applied and the quantitative assessments exposed demonstrate the superiority in relation to other approaches present in the literature.

The paper of Vos et al. (2019) presents a proposal for unsupervised registration as opposed to supervised techniques. Supervised techniques demand pre-defined registration examples and obtaining such examples is not a simple task. Therefore, the authors

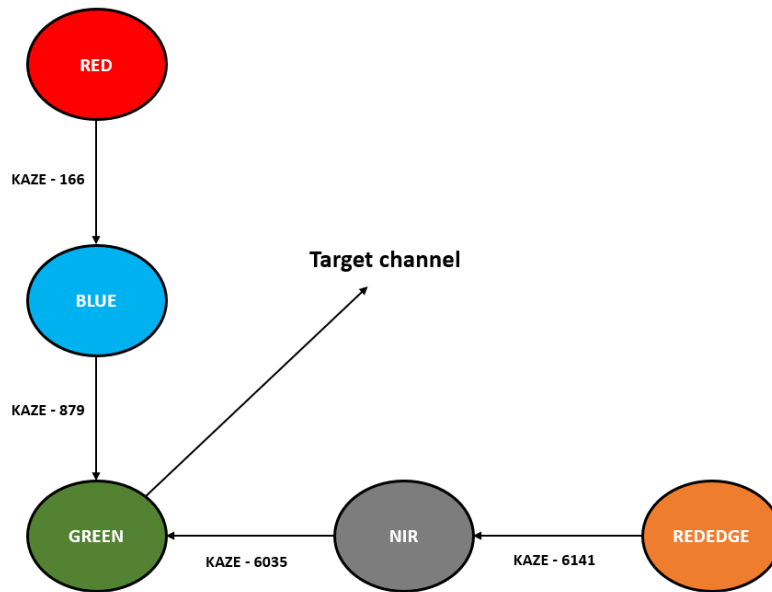


Figure 18 – Framework proposed by Dias Junior et al. (2020) and adapted from Yasir et al. (2018).

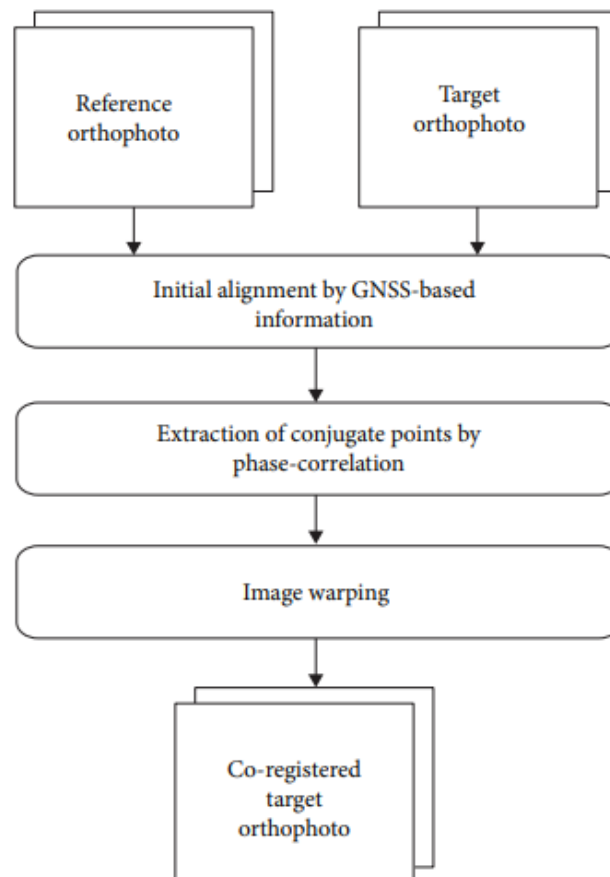


Figure 19 – The flowchart of the proposed co-registration approach by Han et al. (2019).

proposed a structure, called Deep Learning for Image Registration (DLIR), to register deformed images in an unsupervised manner.

The proposed structure is trained to perform image registration through image similarity to the conventional image registration (based on intensity). Later, this training can be used to register pairs of images not yet presented to DLIR. Also, the authors proposed flexible designs for affine registration and elastic registration.

Deformation recognition in UAVs' images using deep learning

During the flight, the UAV is subject to several environmental conditions, which can generate deformations that need to be corrected later. Thus, previous identification of the deformation can help the task of correct them. So, in this chapter, we will discuss the ability of CNN to identify and classify linear and non-linear deformations in images captured by UAV in PA.

4.1 Image Datasets

To analyze the ability of CNN to identify the presence of deformations in an image captured by a UAV in PA, we consider some linear (rotation, translation, and perspective) and non-linear deformations in our experiments. These images were acquired using an eBee SenseFly mapping drone. The senseFly S.O.D.A. camera of 1in Sensor was used with RGB lens F/2.8-11, 10.6 mm. Under these conditions, the Ground Sample Distance (GSD) is $5\text{cm}/\text{pixel}$ (SILVA et al., 2020b).

From two mosaics (both sugar cane, but from different areas), we selected patches to compose our datasets. These mosaics are RGB, but for conducting our experiments we convert gray scales. In addition, both mosaics were 18543×2635 and 8449×11180 pixels size and the patches obtained were 128×128 pixels size. These datasets will be called Data Set 1 (DS1) and Data Set 2 (DS2) respectively. Figure 20 illustrates how we extract the patches from the mosaics.

Posteriorly, we discard patches that have non significant visual information. This was determined by the number black pixels (n) in the patch. Thus, if $n < 10$, the patch is considered for the composition of the dataset; otherwise, the patch is discarded. Therefore, we built two datasets, which have, respectively, 3353 and 2365 images. Figure 21 shows examples of two patches, each obtained from a different mosaic.

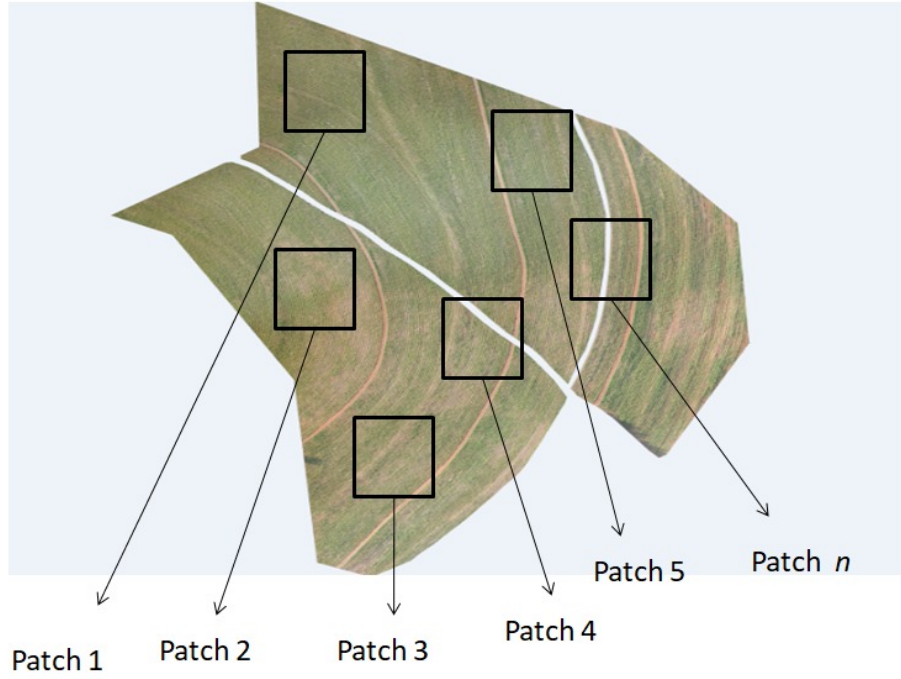


Figure 20 – For the construction of both datasets, we use the same strategy, which consists of making cutouts in a mosaic to extract the patches.

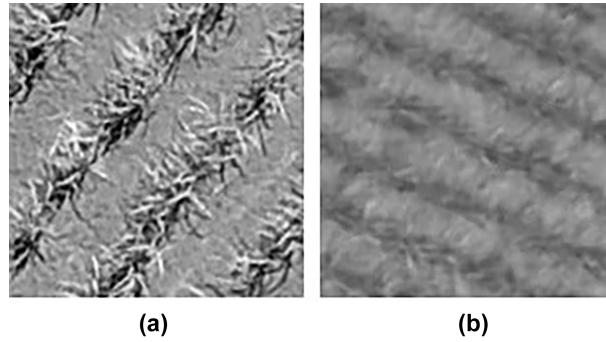


Figure 21 – Example of patches: (a) an patch of DS1; (b) an patch of DS2.

4.1.1 Linear deformation dataset

For linear deformations, we consider three types commonly observed during the flight of a UAV: rotation, translation, and perspective. Subsequently, we applied some variations of the linear transformations to the patches, according to the following parameters:

- ❑ **Rotation:** we used $\theta = \{0^\circ, 5^\circ, 10^\circ, 15^\circ\}$, thus generating 4 classes of rotated patterns.
- ❑ **Translation:** patches were translated by 25 pixels in 4 possible directions: right and top; right and down; left and top; and left and down, and the original patch, thus generating 5 classes.

- **Perspective transformation:** it simulates UAV up and down possibilities in moments of patch capture (see Figure 23). We keep the proportion of lines identical to the original patch. For the columns, the proportions in each of the distorted classes were: (0.05, 0.66); (0.05, 0.77); (0.02, 0.66); (0.02, 0.77).

It is important to mention that these deformations results in black areas, especially at the edges of the patch, which can influence the training and testing of CNN. So, to remove these areas, we cut the 64×64 pixels region aligned with the center of the patch, thus removing any artifact added to the patch by the selected transformation method, as shown in the Figure 22.

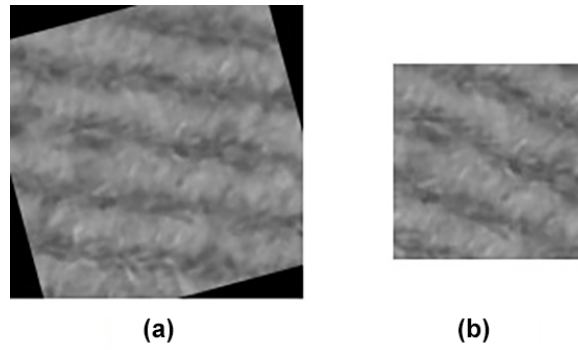


Figure 22 – (a) Patch after a 15-degree rotation transformation. Notice that this patch presents black areas which can directly influence the neural network training and testing; (b) Cropped region with 64×64 pixels size.

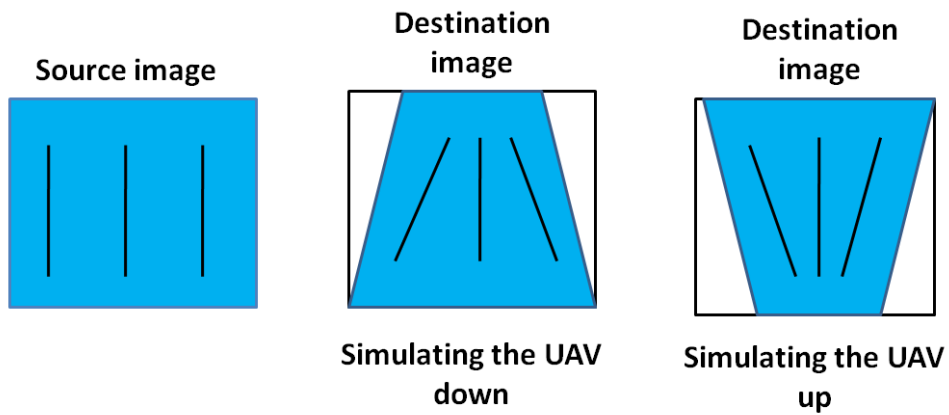


Figure 23 – Perspective transformation in UAV up and down simulations.

We trained the network with the subtraction of the deformed patches and their deformation-free patches, to relate them. Thus, all artificially distorted patches underwent a subtraction operation from the original patch. Let A be the distortion-free patch

and B the distorted patch relative to A , we define X as the patch resulting from the subtraction operation and to be processed by the CNN. The operation performed between A and B is defined pixel by pixel. We must also consider that the subtraction operation may result in negative values and an patch is expected to have only positive values. To avoid that, we normalized the computed x_{ij} values as follows:

$$x_{ij} = \max(b_{ij} - a_{ij}, 0) \quad (1)$$

where $a_{ij} \in A$ represents a pixel of patch A , $b_{ij} \in B$ represents a pixel of patch B and $x_{ij} \in X$ represents a pixel of patch X . Figure 24 illustrates this between an artificially distorted and a distortion-free patch.

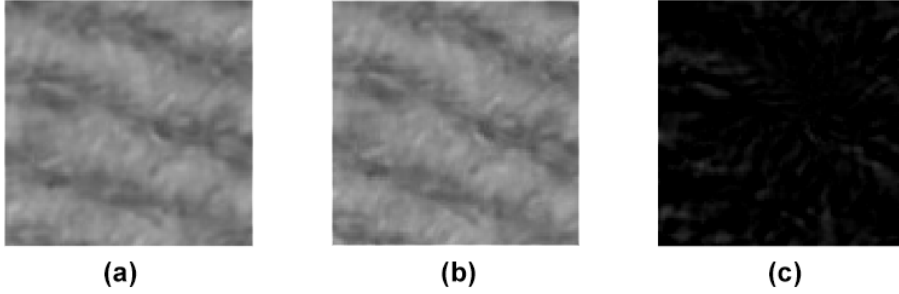


Figure 24 – Example of subtraction operation between two patches: (a) Artificially distorted patch; (b) Corresponding distortion-free patch; (c) Result from the subtraction operation with (a) and (b) by Equation 1.

4.1.2 Non-linear deformation dataset

Literature presents many works addressing the problem of non-linear deformations, in a varied contexts (WALIMBE; SHEKHAR, 2006; SHEKHAR; ZAGRODSKY, 2002; WANG; STAIB, 1998). A Eulerian Reference can be used in the formulation of the non-linear model. For this scenario, a component of the patch is tracked in relation to its final coordinates (WANG; STAIB, 1998). For that, we define the transformation through a homeomorphic mapping of the coordinate system in two dimensions, according to Equation 2:

$$w = (z, y) \rightarrow (x - u_x(w), y - u_y(w)) \quad (2)$$

where $u(w) = [u_x(w), u_y(w)]^T$ is the displacement at each pixel w whose coordinate is denoted as (x, y) (WANG; STAIB, 1998).

The formulation of the non-linear model is also defined in Christensen et al. (1996), where components are considered to be displaced with a proportional force. This spatial transformation satisfies the Partial Differential Equation (PDE), as shown in Equation 3:

$$\mu \nabla^2 u + (\mu + \beta) \nabla(\nabla \cdot u) = F(u) \quad (3)$$

where $\mu(w) = 0$ for w on the image boundary. In Equation 3, μ and β are *Lamé* constants. The body force, $F(u)$, is the driving function that deforms the images.

To generate non-linear deformations for our experiments, we consider what was proposed by Eppenhof et al. (2019). Thus, to create these deformations in the patches, deformable transformations are inserted as B-spline transformations and the displacements are defined by a random grid. In Eppenhof e Pluim (2019), as we deal with two-dimensional images, we use two grids, one for the displacements in the *y-direction*, and one for the displacements in the *x-direction*. Figure 25 shows the relationship of an patch in our dataset with its respective grid and the distortion generated in that patch through the grid.

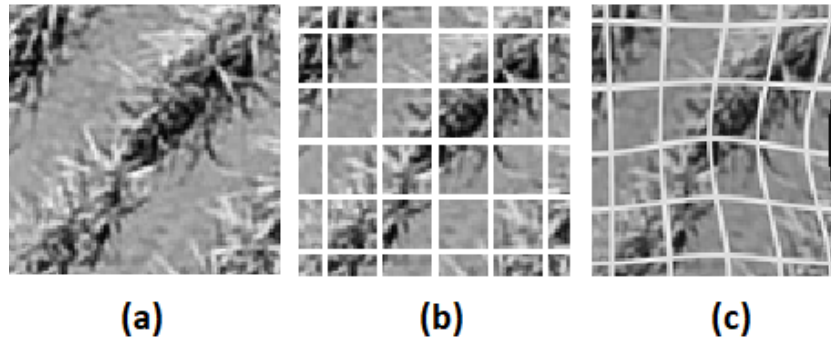


Figure 25 – Example of non-linear distortion: (a) Original patch; (b) Original patch (a) with their respective grids; (c) Patch after applying the deformation.

We also concatenate two rigid deformations in the non-linear deformed patch using an interpolated method. These transformations applied to the patch grid are rotation and scale, where the rotation varies from 0 to 45 degrees and the scale at $\pm 10\%$, always around the point $(0.5, 0.5)$. Figure 26 presents an example of the result of concatenating non-linear and linear transformations.

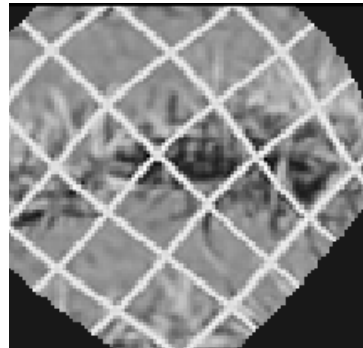


Figure 26 – Combination of transformations for the patch in Figure 25 (c).

As we did previously for patches with linear deformations, here we also cropped 64×64 pixels size from the center so that these black areas do not influence the subsequent training and testing of the network. We also need to mention that the deformation-free

patches were also cropped under the same conditions. Figure 27 shows an example of a cropped patch.

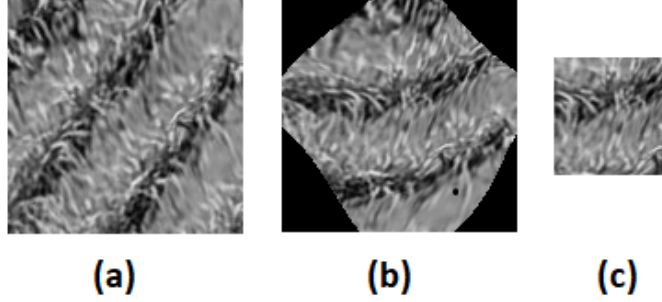


Figure 27 – (a) Deformation-free patch; (b) Deformed patch, according to (EPPENHOF; PLUIM, 2019); (c) Cropped patch.

4.2 Experiments

In order to classify and consequently recognize the presence of deformations in an patch mentioned in sections 4.1.1 and 4.1.2, we consider 4 traditional CNN architectures: InceptionV3 (Szegedy et al., 2016), ResNet (He et al., 2016), SqueezeNet (IANDOLA et al., 2016) and VGG-16 (SIMONYAN; ZISSERMAN, 2014). We used these networks pre-trained in the 2012 dataset Imagenet and made the pertinent adjustments (e.g., input/output sizes) to our classification problem (SILVA et al., 2020a; SILVA et al., 2021).

Motivated by Marcos, Rodovalho e Backes (2019a) and Marcos, Rodovalho e Backes (2019b), where simpler CNN and sets of filters were used to solve less complex classification problems, we proposed a CNN architecture to solve the problem in question. Our CNN is inspired and, in general, consists of the simplification of AlexNet. We used 5 convolutional layers to process the patches and each layer has, respectively, 16, 32, 32, 64, and 128 filters of size 3×3 . To improve its learning ability and to enhance its training, we applied the ReLU non-linearity activation function after each convolutional layer. There is also a batch normalization after each ReLU function, which is followed by a 2×2 max-pooling layer. In the sequence, we used the resulting volume as an input for the dense layers. The first two dense layers have 128 neurons each and they use ReLU as activation function. We also apply a 20% dropout in each dense layer. Finally, the output layer has 2, 4, or 5 neurons that determine the equivalence class. Note that the output layer is related to the analyzed deformation, as explained in the previous section. The rotation considers 4 classes of equivalence, translation and perspective consider 5 classes, and the non-linear ones consider two classes (presence or not). Figure 28 shows the structure of our CNN.

For our experiments, we divided DS1 and DS2 into two sets: training and validation. The training set had 75% of the samples, while the remaning patches were used for validation.. Each deformation problem was trained in 20 epochs, in order to maximize

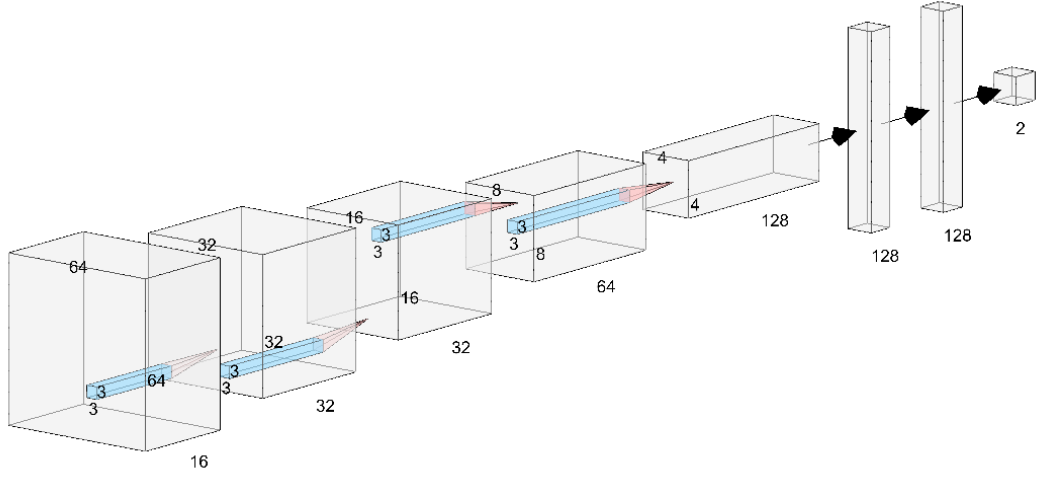


Figure 28 – Illustration of the architecture of our CNN.

accuracy. Experiments were conducted on a Personal Computer with Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz, 32GB RAM, 64-bit Windows OS and GPU NVIDIA GeForce GTX 1050 Ti, 4GB GDDR5. We also used Python 3.6 and Keras 2.1.6-tf with TensorFlow 1.10.0 and CUDA Toolkit 9.0 to implement and test the experiments (ABADI et al., 2016).

4.3 Results and Discussions

The results of our experiments demonstrate that our CNN performs better than traditional ones when we consider linear deformations. Table 1 summarizes the precision results of all the CNN analyzed in relation to rotation, translation, and perspective. We notice that the best performance is obtained when dealing with the problem of patch rotation. For this problem, our CNN model is capable to classify the rotation distortion with 99.85% and 99.18% accuracy in the DS1 and DS2 datasets, respectively. Moreover, the CNN presents a good ability to generalize the features learned in the training set to the test set. This may be explained by the fact that the rotation operation results in a less distorted patch in comparison to other patch transformations, i.e., an easier classification problem.

For perspective problem, our CNN model achieves high accuracy, especially for DS1, which reached 95.50%. For the same problem, however, we obtained only 91.47% for DS2 accuracy. Notice that this accuracy is substantially lower when compared to the rotation experiment. One explanation for this behavior is that this kind of transformation affects differently the regions of the sample, while rotation affects all points of the sample equally. Moreover, although both datasets present a lower result, we notice that dataset DS1 presents a superior result when compared to dataset DS2. It may be the case that the number of samples in the training set, which is larger in DS1, contributes positively to learn this transformation.

For the translation transformation, we noticed a considerable drop in the accuracy of the network when evaluating the DS2 dataset (70.24%). Even though the dataset DS1 also present an inferior performance when compared to the rotation transformation, its result is superior to the ones obtained for the perspective transformation. This result observed in the DS2 dataset is probably explained by the lack of details in their original patches, as illustrated in Figure 21. Since crop lines and land regions present similar gray-level distributions, the result of the subtraction operation between the original and the translated patch results in a mostly black patch, i.e., an patch without enough attributes for our CNN to learn.

Regarding non-linear deformations, our CNN achieved an accuracy greater than 92%, however, it has a slightly lower performance than traditional architectures, as shown in Table 2. On the other hand, our CNN has a considerably lower computational cost compared to traditional ones. Table 3 shows the number of parameters trained in each CNN in the context of our problem. When compared with InceptionV3, our CNN model obtained a accuracy 1.72% and 5.92% lower in dataset DS1 and DS2, respectively. However, its computational cost is smaller as it has only 0.64% of the number of parameters of the InceptionV3 model. Additionally, we must emphasize that compared CNNs were previously trained in the 2012 Imagenet dataset while our proposed model was trained from scratch. In this sense, given that the results of our architecture does not differ from the traditional CNN, this approach may enable a great advance for the use of this resource in a system embedded in the UAV. The UAV flight time detection can support subsequent processes (e.g., correction of detected deformation and image registration), in addition to reducing the financial costs inherent in the process.

Table 1 – Results (%) obtained for our CNN and the compared ones for linear deformations.

CNN model	Translation		Rotation		Perspective	
	DS1	DS2	DS1	DS2	DS1	DS2
ResNet	91.83	48.13	95.00	96.84	59.55	62.63
InceptionV3	20.00	60.10	98.48	98.23	20.00	65.96
VGG-16	94.76	65.15	98.63	98.74	84.89	75.20
SqueezeNet	90.51	40.40	91.77	96.15	55.68	55.20
Proposed	96.92	70.24	99.85	99.18	95.50	91.47

The results show the ability of CNNs to correctly identify possible deformations inherent to the flight of a UAV, whether these deformations are of a linear or non-linear nature. Also, an architecture with lower computational cost, but with still high accuracy, proved to be effective in treating the problem.

The proposed architecture can eventually be used in a system embedded in the UAV, which could already detect and mark a certain image that has suffered some type of

Table 2 – Results (%) obtained for each CNN model for non-linear deformations.

CNN model	DS1	DS2
ResNet	98.95	94.36
InceptionV3	99.25	98.84
VGG-16	98.88	95.35
SqueezeNet	98.43	96.41
Proposed CNN	97.53	92.92

Table 3 – Number of parameters of each CNN model.

CNN model	Number of parameters
ResNet	22,591,810
InceptionV3	22,081,826
VGG-16	14,797,122
SqueezeNet	723,522
Proposed	141,058

deformation. This information would be extremely valid for the subsequent deformation correction process.

Deformable co-registration using deep learning

In this chapter, we present an approach based on deep learning for multispectral images registration obtained by UAVs in PA. Our proposal is inspired by (EPPENHOF et al., 2019), where a U-Net was adapted to estimate the vector deformation field of a moving image in relation to the same fixed image.

For our experiments, we consider two multispectral datasets from different crops planted. We include linear and non-linear deformations artificially in the bands of these datasets and, later, we evaluate the ability of networks to estimate the field and perform the registration between the bands of a given scene.

5.1 Image Dataset

Our experiments considered two dataset to evaluate the approach proposed. Both datasets have 1280×960 pixels size image and an average of 80% overlap between images. The spectra present in the images are Blue, Green, Red, Near-Infrared (NIR), and Red Edge. In each dataset, the images were obtained in a single flight, without any type of pre-processing. This flight took place at an average height of 100 meters, at an average speed of $20m/s$. Under these conditions, the GSD is $6.8cm/pixel$ (JUNIOR et al., 2020).

Due to the nature and purpose of the UAV flight, both dataset have images of heterogeneous content. Figure 29 show image samples of the soybean dataset and Figure 30 shows image samples of the cotton dataset.

5.2 Data preparation

Subsequently, for conducting and evaluating the experiments, the images were manually marked by two different experts, to have a ground truth, where the bands were previously aligned. Each expert marked a different dataset. In general, the experts

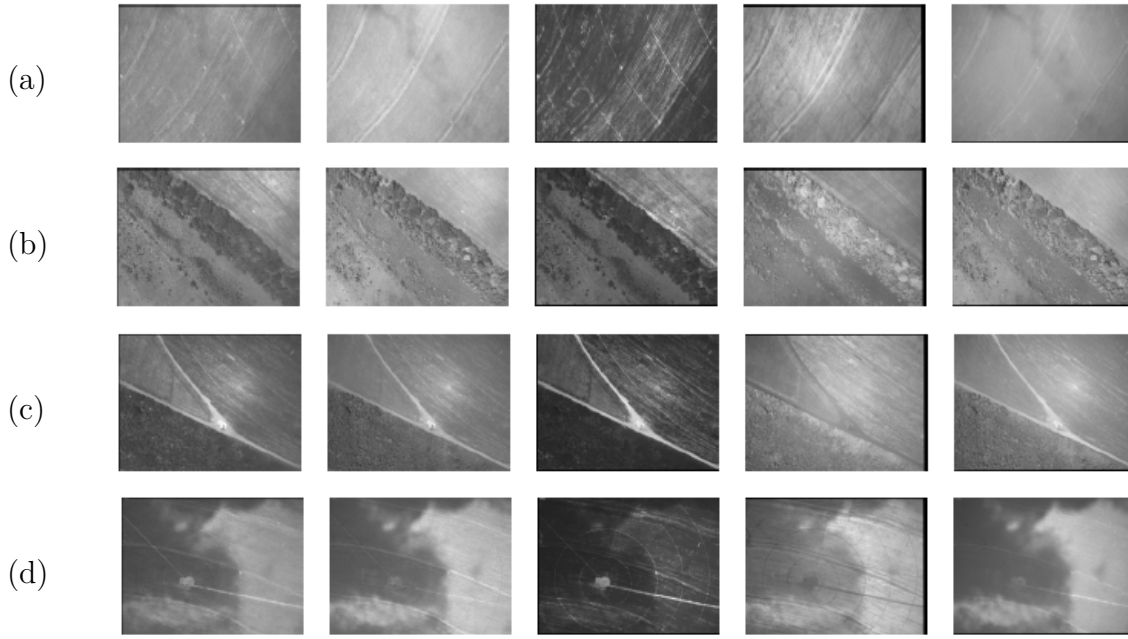


Figure 29 – Examples of images from soybean dataset: (a), (b), (c), and (d) present, respectively, four different scenes captured by UAV, with the columns represent the respective bands (Blue, Green, Red, NIR, and Red Edge, respectively) of the scene.

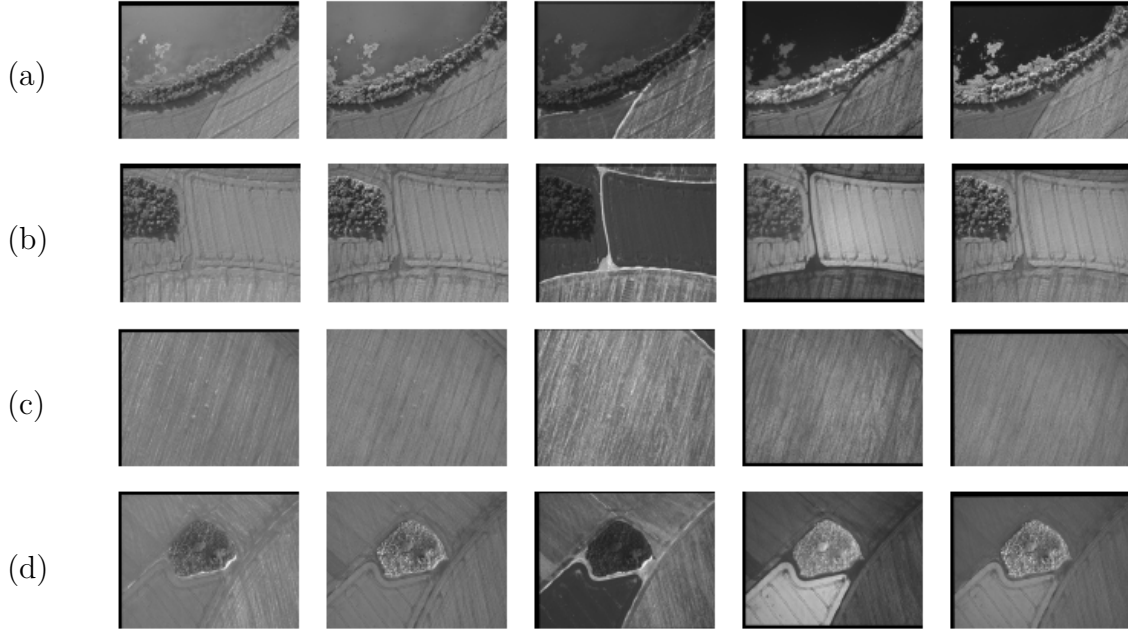


Figure 30 – Examples of images from cotton dataset: (a), (b), (c), and (d) present, respectively, four different scenes captured by UAV, with the columns represent the respective bands (Blue, Green, Red, NIR, and Red Edge, respectively) of the scene.

marked 12 points in the green band. In the sequence, the experts marked the equivalent points in the other bands (JUNIOR et al., 2020). Tables 4 and 5 shows the misalignment,

in pixels, between the ground truth generated by the experts in the soybean and cotton datasets respectively.

Table 4 – Misalignment average, in pixels, between the sensors present in soybean dataset.

	Blue	Green	Red	NIR	Red Edge
Blue	-	5.18	15.75	15.07	12.14
Green	5.18	-	15.09	12.33	4.02
Red	15.75	15.09	-	29.25	14.79
NIR	15.07	12.33	29.25	-	16.17
Red Edge	12.14	4.02	14.79	16.17	-

Table 5 – Misalignment average, in pixels, between the sensors present in cotton dataset.

	Blue	Green	Red	NIR	Red Edge
Blue	-	28.30	12.11	33.28	8.48
Green	28.30	-	21.44	24.01	35.66
Red	12.11	21.44	-	14.94	21.30
NIR	33.28	24.01	14.94	-	39.37
Red Edge	8.48	35.66	21.30	39.37	-

To perform our experiments, we have proportionally reduced the images to 20% of their original size, preserving their content and proportionality. Thus, with this resizing, we do not insert any deformation. The purpose of this operation is to reduce the computational cost of further training in relation to the original size of the images. In this way, the images will be 256×192 pixels size.

For the experiments, according to Junior et al. (2020), we consider the green band as fixed. Junior et al. (2020) justifies the choice of green as the band to be considered fixed because the images are mostly composed of vegetation, so this band has more content that can be used as the reference object for the matching of the other bands. In this way, we generate deformations in the other bands and we will keep the green preserved for both datasets.

The non-linear deformations consisted of two random grids, one for displacements in the *y-direction*, and the other for displacements in the *x-direction*. In short, the transformation is defined through a 3×3 grid point B-spline grid, where random displacements are ranging from a uniform distribution. Table 6 shows the parameters mentioned for the deformation (EPPENHOF; PLUIM, 2019).

Table 6 – Parameters for B-spline transformations. Vector displacements are sampled uniformly in the distributions in the given ranges.

Grid point displacement		
Grid size	x	y
3×3	$[-0.05, 0.05]$	$[-0.05, 0.05]$

When applying the non-linear transformation, we can see the result in the moving image, as shown in Figure 31. Note that the grid in this figure is for a better visual representation of the deformation and does not represent in size the B-spline transformation grid.

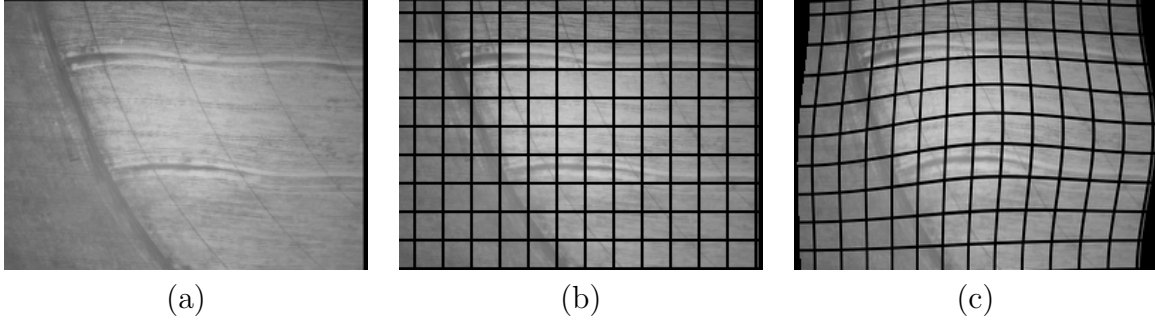


Figure 31 – Example of distortion in near-infrared band: (a) original image; (b) original image (a) with their respective visual grids; (c) image after applying the deformation.

To make the generated deformations even more adherent with real deformations, we also inserted two more rigid deformations often found in an UAV flight: rotation and scale. These deformations were also random in a controlled range of variations. For the scale, we consider variations of $\pm 2\%$, with the original. For the rotation, we consider a variation of up to 10 degrees around the point $(0.5, 0.5)$. Figure 32 shows the same image as in Figure 31-(c) with the insertion of the linear deformations.

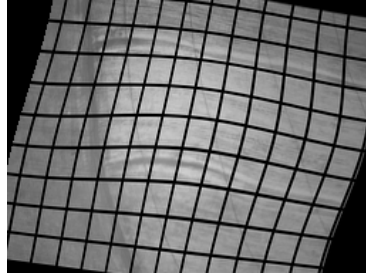


Figure 32 – Combination of two linear transformations (rotation and scale) for the image in Figure 31-(c).

Our approach will be based on supervised learning. The aim will be to learn the displacement field (U) of the moving image to the fixed image. The displacement field (U) is calculated as follows:

$$U = originalGrid - transformedGrid \quad (4)$$

Where *originalGrid* refers to the deformation-free image grid (original image) and *transformedGrid* refers to the image grid after inserting the deformations. The visual pre-

sensation of the displacement fields, with the respective color map, is shown in Figure 33.

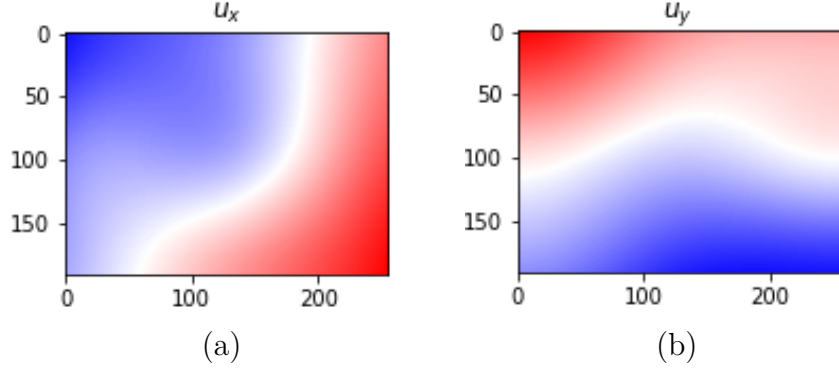


Figure 33 – Example of displacement fields in Figure 32 in relation to Figure 31-(b): in (a) we have the displacement field for the x axis; in (b) we have the displacement field for the y axis. The colors blue and red represent positive and negative displacements respectively.

5.3 End-to-end training

Inspired by (EPPENHOF et al., 2019), the network is trained to maximize the accuracy of the estimate between the estimated vector field ($\hat{U}$) and the true vector field (U). In addition, the loss is defined as:

$$loss = \frac{1}{n} \sum_{j=1}^n |u_j - \hat{u}_j| \quad (5)$$

where n refers to number of iterations. We also use the Adam optimizer with a learning rate in 10^{-4} .

The end-to-end architecture used is the fully-convolutional neural network proposed by Ronneberger, Fischer e Brox (2015). This architecture is known as U-Net and its use is mainly focused on the segmentation of 2D and 3D medical images (HEO et al., 2019). According to (EPPENHOF et al., 2019), to adapt the U-Net to the registration problem, the input layer has two channels, one considered the fixed image and the other for the moved image. The output layer is also adapted to have two channels, where each channel corresponds to the form vector field on one of the axes (x and y). Finally, all activation functions are the Leaky ReLU parameterized as $\phi(x) = \max(x, 0.01x)$, which restrict the “dying” of neurons that occurred using ReLU (EPPENHOF; PLUIM, 2019). However, the output layer has no activation function to enable real values. Figure 34 shows the proposed architecture.

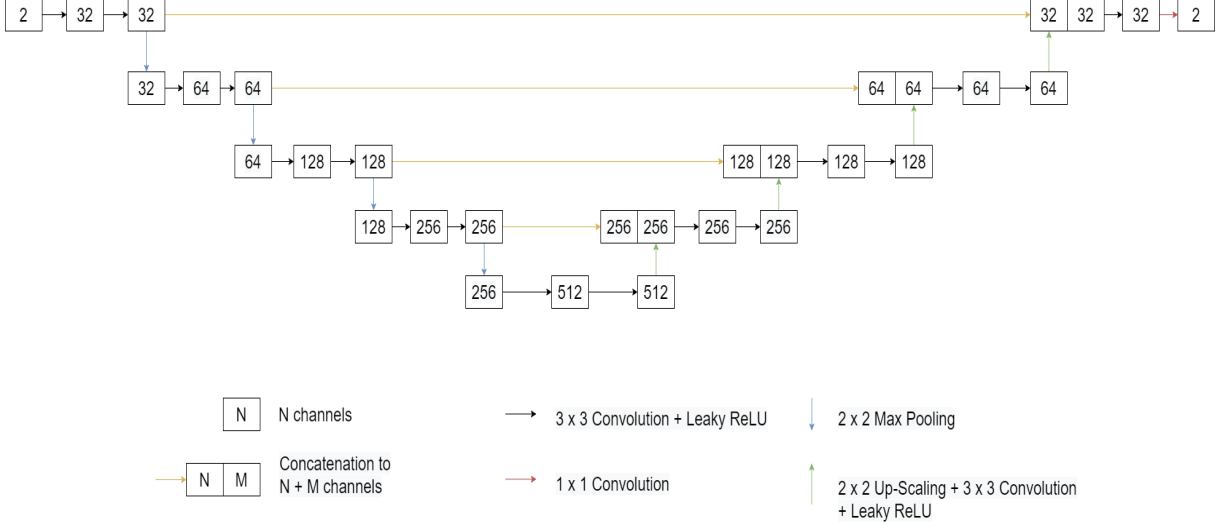


Figure 34 – The proposed network architecture. The network takes two bands as input, and outputs two maps: one for each vector field axis. Adapted from (EP-PENHOF; PLUIM, 2019).

5.4 Experimental Results in RGB Dataset

First, to evaluate the ability of our network to predict the vector field of deformation between two images, we tested it on a set of images in gray scales. This set has a mix of DS1 and DS2 images mentioned in the Chapter 4. Each image was deformed under the same conditions presented in section 5.2, however, the images in this set are 128×128 pixels size. The dataset has a total of 328 images and was divided into sets of training, validation, and tests, in the proportion of 70%, 15%, and 15% respectively. Experiments were conducted on a Personal Computer with Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz, 32GB RAM, 64-bit Windows OS and GPU NVIDIA GeForce GTX 1050 Ti, 4GB GDDR5. We also used Python 3.6 and Keras 2.1.6-tf with TensorFlow 1.10.0 and CUDA Toolkit 9.0.

In this scenario, the network input consists of a two-channel image: the first channel contains the deformation-free image, while the second channel contains the deformed image. The training was carried out with 500 epochs and obtained an accuracy of **97.54%**. Figure 3 summarizes the network’s performance for this scenario in terms of accuracy and loss.

With the vector fields predicted by the network, we need to map the points of the deformed image towards the fixed image. In short, we apply the transformation learned by the network, interpolating them in a new grid, as follows:

$$I_R(x, y) = I_M(x + U(x, y), y + V(x, y)) \quad (6)$$

where I_R is the registered band, I_M is the moved band and $U(x, y)$ and $V(x, y)$ are the predicted vector fields for the x and y axes respectively.

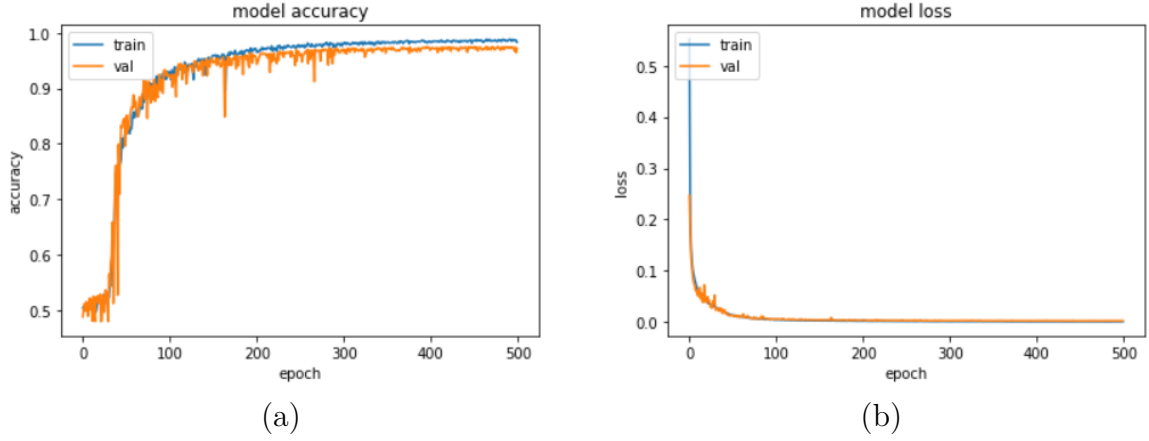


Figure 35 – Model accuracy in RGB dataset: In (a) we have a model accuracy, and in (b) we have a model loss.

With the displacement field predicted by the network applied to the deformed image, we can visualize the result through an overlap between the two images. Figure 36 shows the overlaps before and after registration. Note that the regions of the overlap refer to when there is the same intensity between the images, while the regions in magenta and green show when the intensities are different.

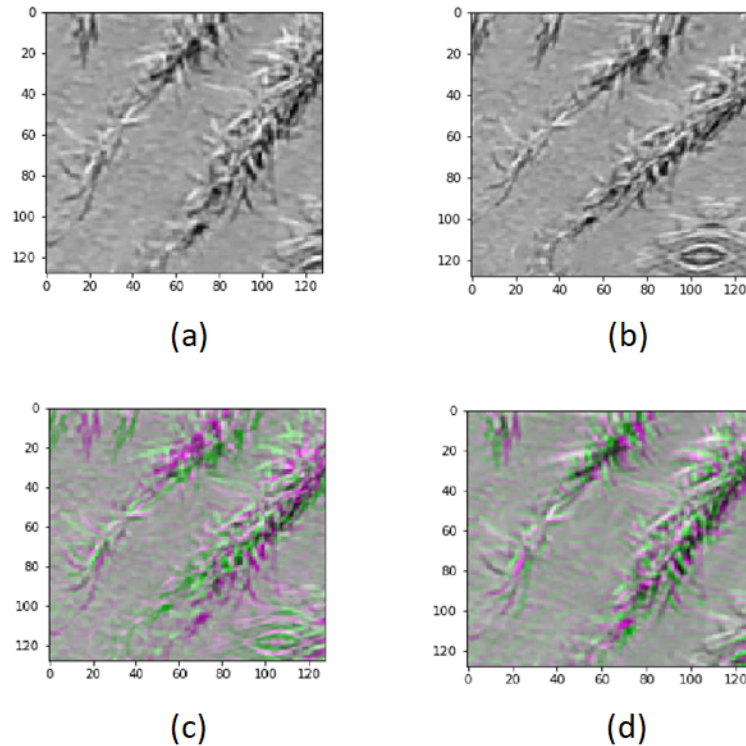


Figure 36 – Example of registration in RGB image from the field predicted by our network: in (a) we have the image free of deformation; in (b) we have the image (a) deformed; in (c) we have the overlap of the images (a) and (b); and in (d) the overlap of the image (a) with the image (b) after the application of the field predicted by the network in (b).

5.5 Experimental Results in Multispectral Datasets

As our network proved to be able to learn the displacement field in gray scales images, we can evaluate the approach proposed in multispectral datasets (soybean and cotton), presented in section 5.2.

In this way, the input of our network consists of an image of two channels. The first channel will always be the fixed band, in our case green. The second channel will be another band, which we want to align towards the green band. In short, we will have the following combinations: (1) Green and Blue, (2) Green and Red, (3) Green and Red Edge, and (4) Green and NIR. The network output, as mentioned, will be the displacement field of the band moved in relation to the fixed band. Figure 37 presents the proposed approach.

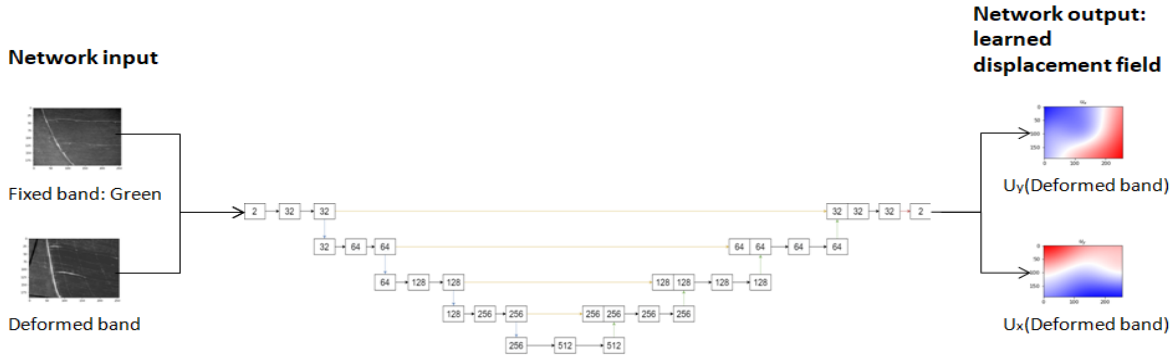


Figure 37 – The proposed approach: the input consists of an image with two channels (a fixed band and another moved band), and the output consists of two displacement fields (in x and y) of the band moved to the fixed band.

Both datasets were divided into sets of training, validation, and test, in the proportion of 70%, 15%, and 15% respectively. Experiments were conducted on a Personal Computer with Intel(R) Core(TM) i7-7700 CPU @ 3.60GHz, 32GB RAM, 64-bit Windows OS and GPU NVIDIA GeForce GTX 1050 Ti, 4GB GDDR5. We also used Python 3.6 and Keras 2.1.6-tf with TensorFlow 1.10.0 and CUDA Toolkit 9.0. In this experimental setup, our approach has about 8.5 million parameters.

Our experiments were conducted in a total of 2000 training epochs in both datasets. For the soybean dataset, in each of the bands considered deformed (Blue, Red, NIR, and Red Edge), we obtained an accuracy ranging from 89.90% to 93.79%. For the cotton dataset, with the same deformed bands as the soybean dataset, we obtained an accuracy ranging from 90.01% to 91.21%. Table 7 summarizes the accuracy for each of the datasets. Also, Figures 38 and 39 show the accuracy of the model each epoch for the soybean and cotton datasets, respectively. Figures 40 and 41 show the model loss for soybean and cotton datasets, respectively.

In the test stage, we present a set containing images that were not part of the training and validation sets, to evaluate the accuracy of the proposed approach for the co-

Table 7 – Summary of validation accuracy for each of the moved bands considered.

Deformed Band	Accuracy	
	Soybean Dataset	Cotton Dataset
Blue	89.90%	90.01%
Red	89.95%	90.11%
NIR	90.50%	91.21%
Red Edge	93.79%	90.54%

registration task. Thus, the displacement fields predicted at this stage were applied to the moving band, also according to Equation 6.

Subsequently, to carry out a visual analysis about the capacity of our approach, we made overlaps between the bands before and after the application of our approach. Figures 42, 43, 44, and 45 show this visual analysis for the soybean dataset. Figures 46, 47, 48, and 49 show the same visual analysis, but for the cotton dataset. Note that these figures for visual analysis have a highlight, where it is possible to observe the ability of our approach to align the images in the correct direction.

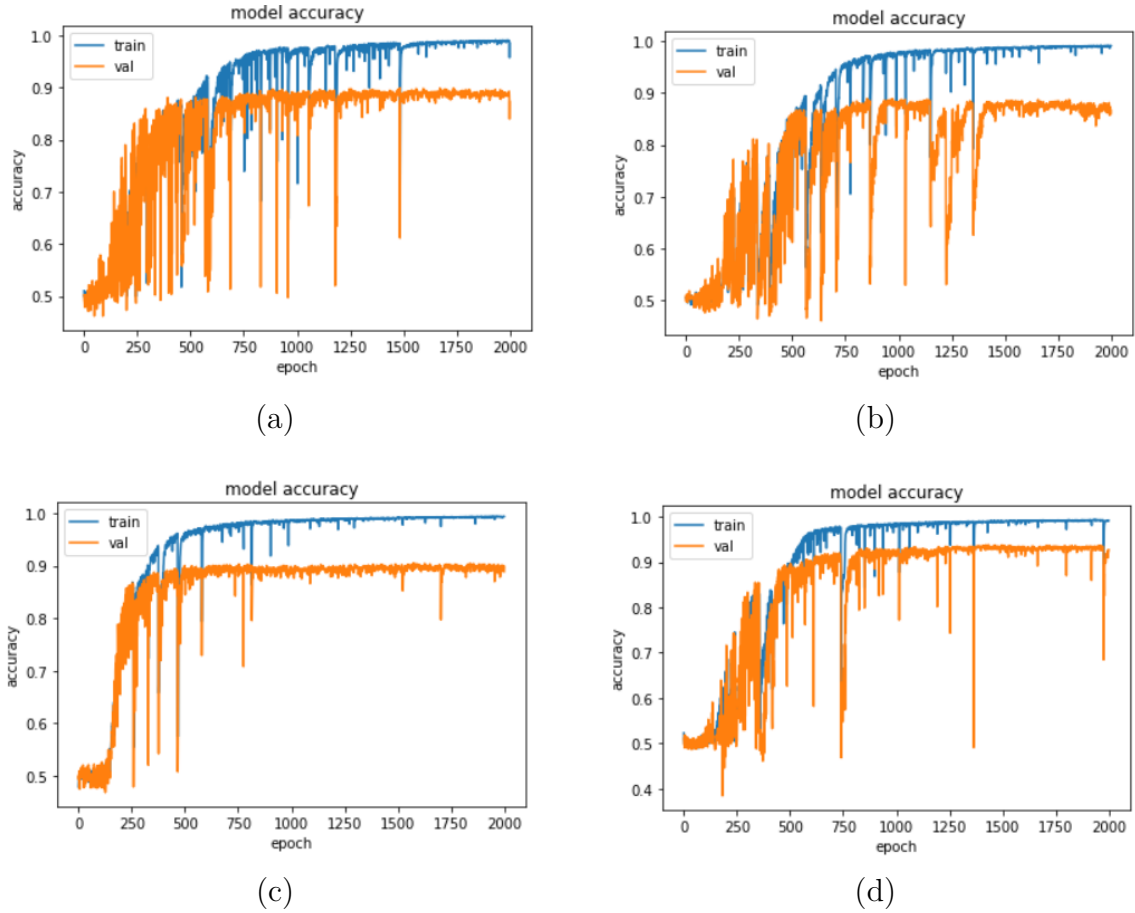


Figure 38 – Model accuracy in soybean dataset: In (a) the Blue band is considered to be moved, in (b) Red, in (c) NIR, and in (d) Red Edge.

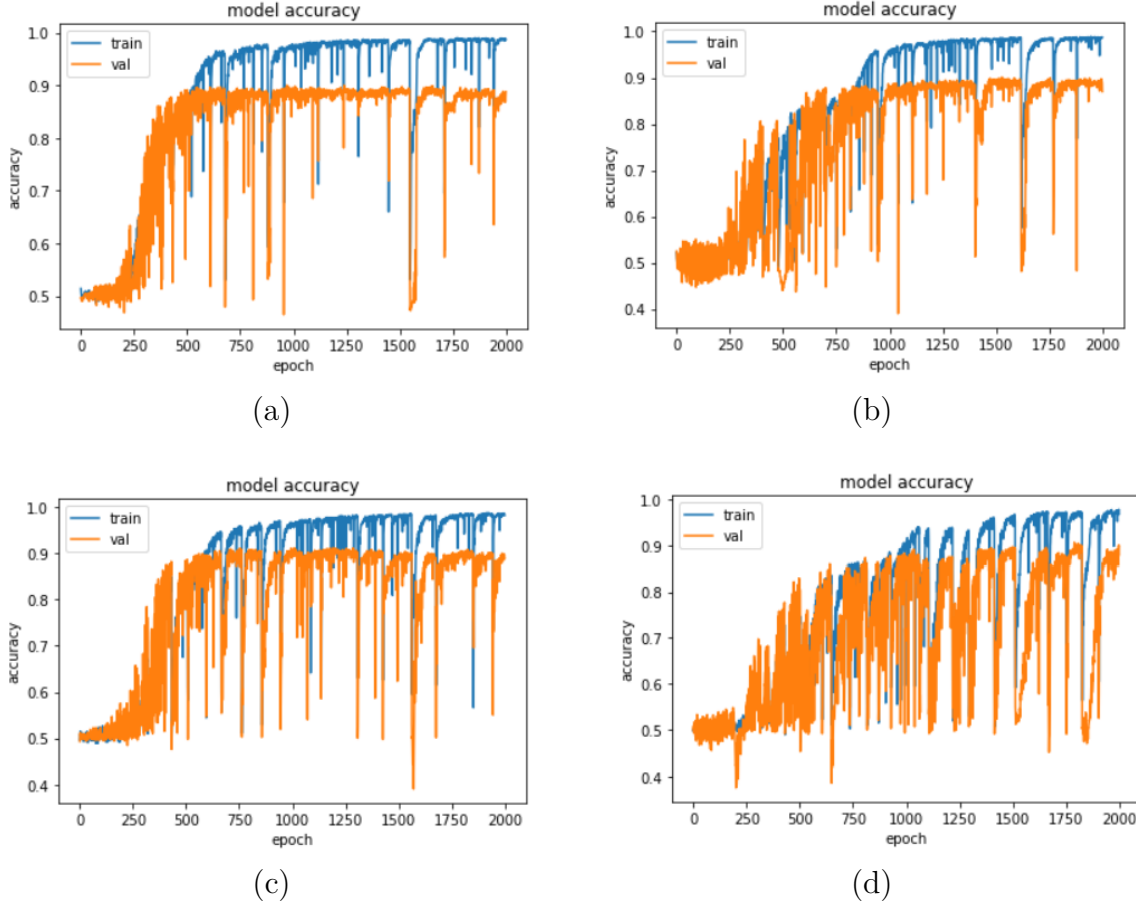


Figure 39 – Model accuracy in cotton dataset: In (a) the Blue band is considered to be moved, in (b) Red, in (c) NIR, and in (d) Red Edge.

5.6 Discussions

The proposed approach proved to be capable of learning a deformation field between bands of multispectral images with considerable precision. The accuracy, when considering both datasets, ranged from 89.90% to 93.79%. In this way, our network is able to perform co-registration after training without the need for the manual marking of points.

In this sense, we need to highlight that although the approach behaved similarly in both datasets, they were obtained under different flight conditions and in different crops, which demonstrates the ability of the approach to generalizing. In addition, the multispectral factor needs to be considered, since as shown in Section 5.4, when we use images in gray scale, our approach demonstrated an accuracy above 97%. The experiments with images in different bands showed an accuracy lower than 97%, probably because the bands carry different information. This is a factor that may cause additional difficulties in our approach to solving the problem.

Although our proposal is related to that of Eppenhof et al. (2019), we take into account multispectral images. In addition to the content of each one of the images in our training set being totally different, in other words, there is no defined pattern of content between

our images (see Figures 29 and 30). In Eppenhof et al. (2019), the images are exclusive of lungs captured through a Computed Tomography (CT).

In Dias Junior et al. (2020), the co-registration between the same bands considered in this work is carried out, however, Dias Junior et al. (2020) does not consider the presence of non-linear deformations, besides needing the manual marking of points for the co-registration task.

Finally, through visual analysis, we can see the ability of our approach to performing co-registration on multispectral images. This capacity is even more evident due to the degree of non-linear deformation considered, in addition to the inclusion of linear deformations, which increases the complexity of learning the network.

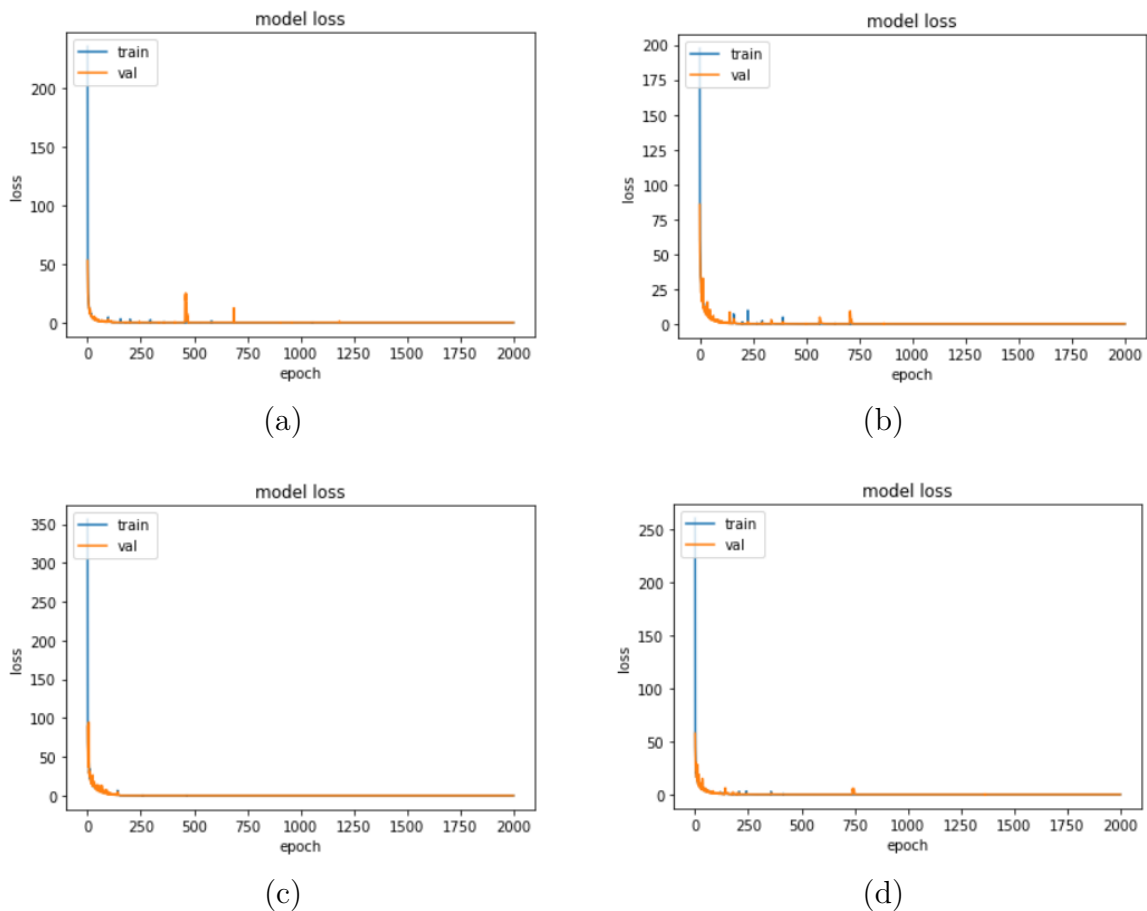


Figure 40 – Model loss in soybean dataset: In (a) the Blue band is considered to be moved, in (b) Red, in (c) NIR, and in (d) Red Edge.

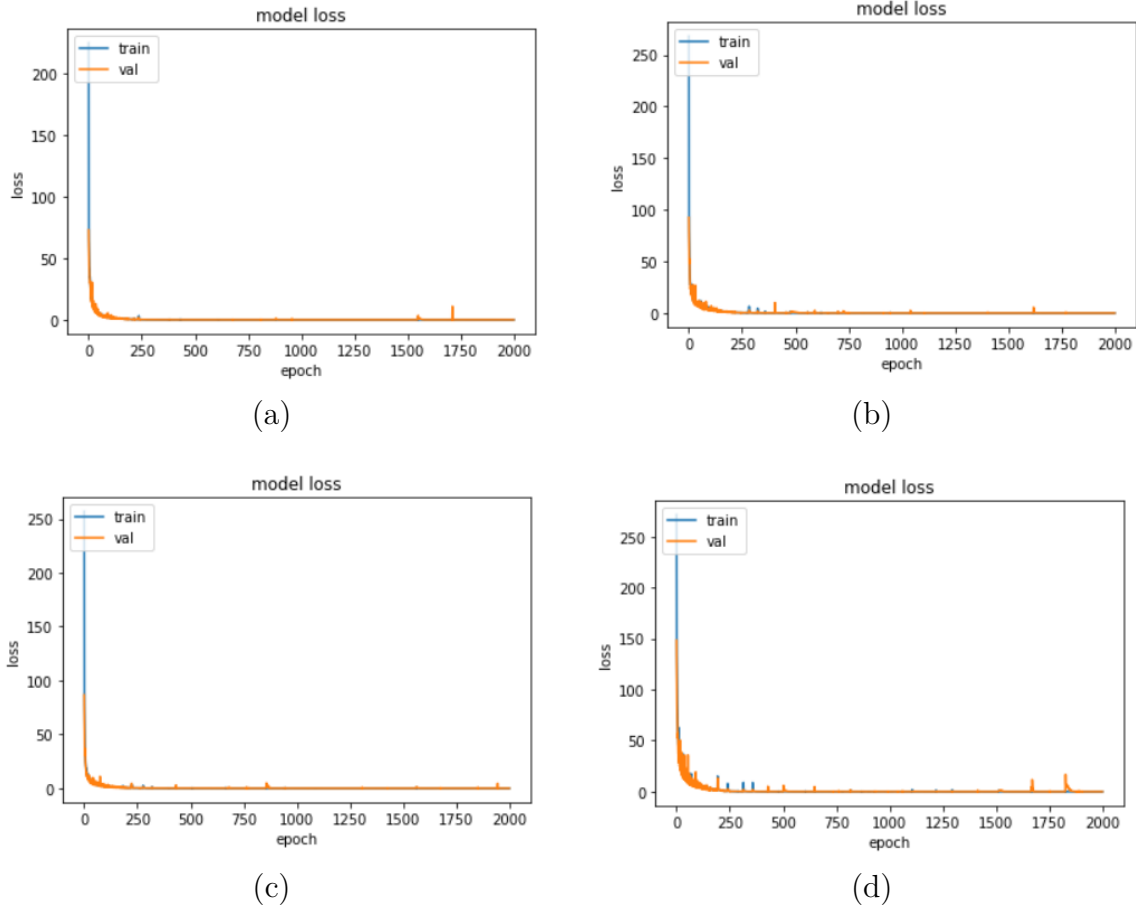


Figure 41 – Model loss in cotton dataset: In (a) the Blue band is considered to be moved, in (b) Red, in (c) NIR, and in (d) Red Edge.

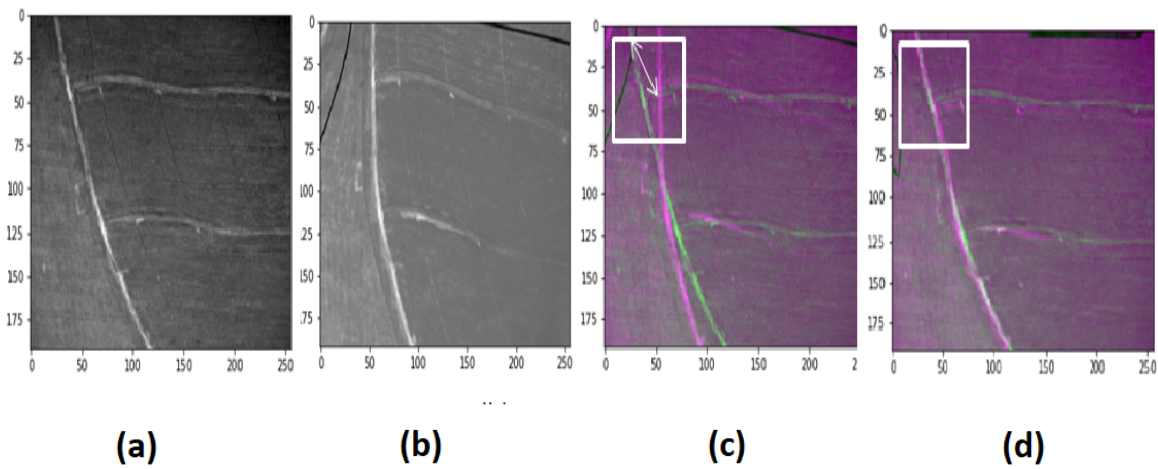


Figure 42 – Example of co-registration in soybean dataset: in (a) we have the fixed band (Green); in (b) we have the moving band (Blue); in (c) we have the overlap of the bands (a) and (b); and in (d) the overlap of the fixed band with the moving band after the application of the field predicted by the network in (b).

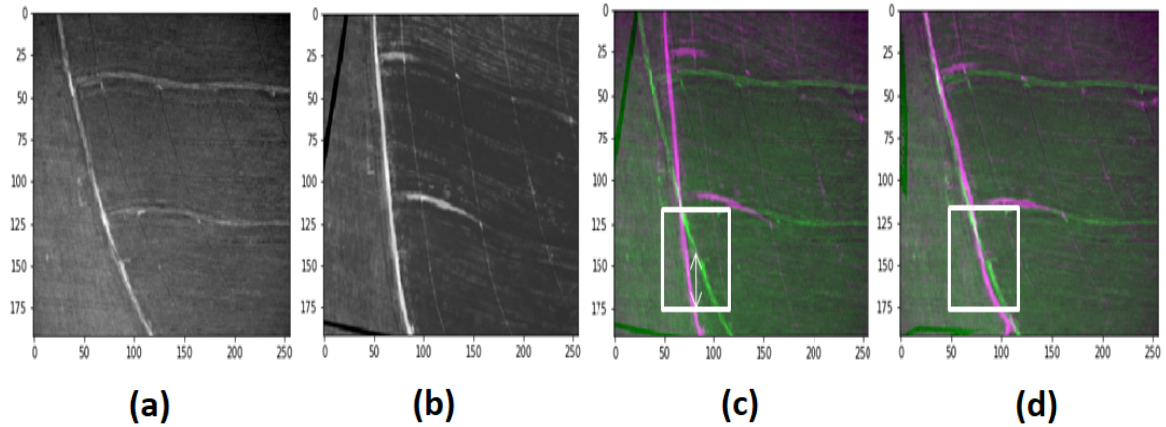


Figure 43 – Example of co-registration in soybean dataset: in (a) we have the fixed band (Green); in (b) we have the moving band (Red); in (c) we have the overlap of the bands (a) and (b); and in (d) the overlap of the fixed band with the moving band after the application of the field predicted by the network in (b).

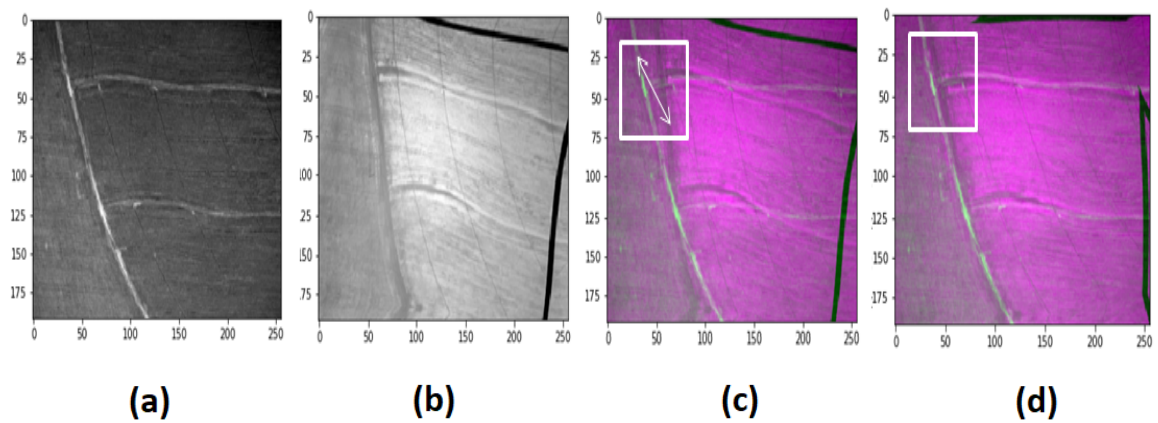


Figure 44 – Example of co-registration in soybean dataset: in (a) we have the fixed band (Green); in (b) we have the moving band (NIR); in (c) we have the overlap of the bands (a) and (b); and in (d) the overlap of the fixed band with the moving band after the application of the field predicted by the network in (b).

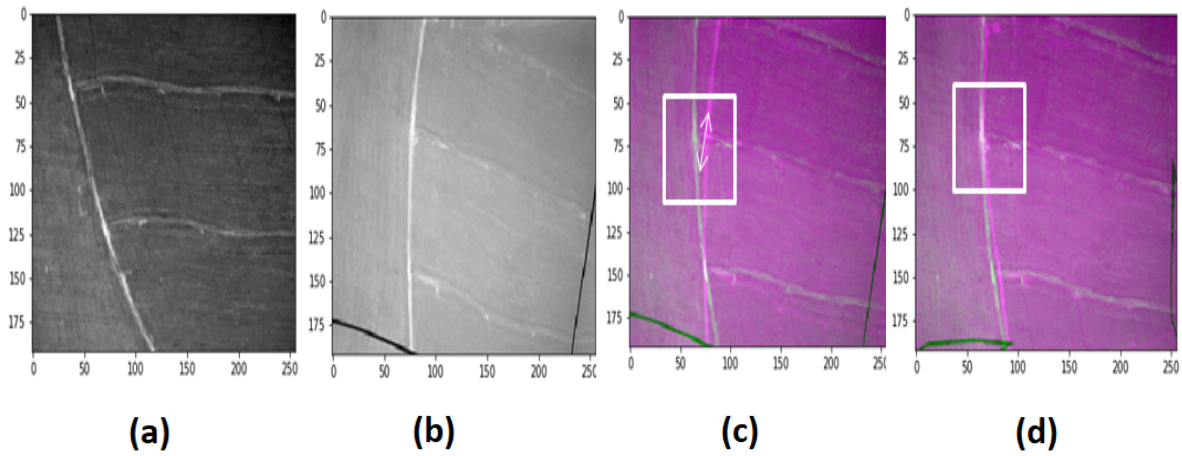


Figure 45 – Example of co-registration in soybean dataset: in (a) we have the fixed band (Green); in (b) we have the moving band (Red Edge); in (c) we have the overlap of the bands (a) and (b); and in (d) the overlap of the fixed band with the moving band after the application of the field predicted by the network in (b).

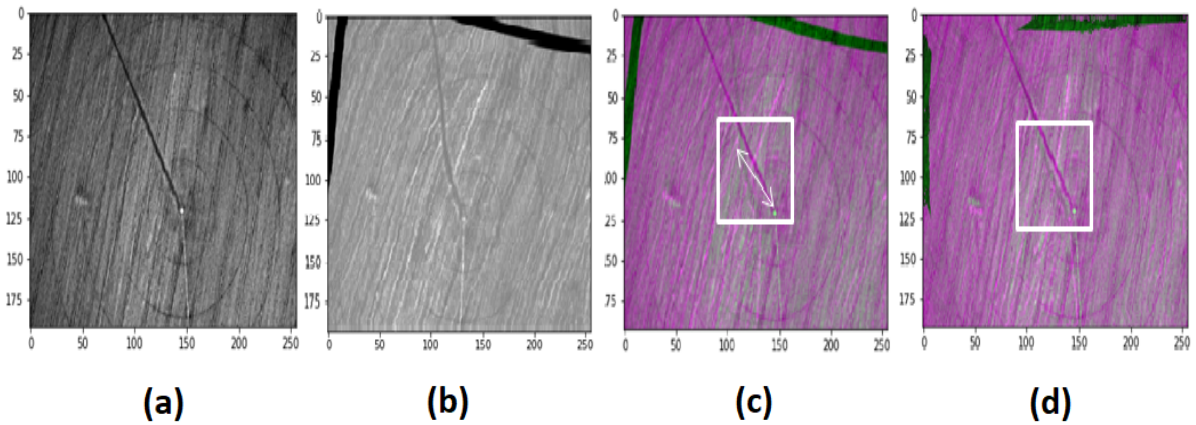


Figure 46 – Example of co-registration in cotton dataset: in (a) we have the fixed band (Green); in (b) we have the moving band (Blue); in (c) we have the overlap of the bands (a) and (b); and in (d) the overlap of the fixed band with the moving band after the application of the field predicted by the network in (b).

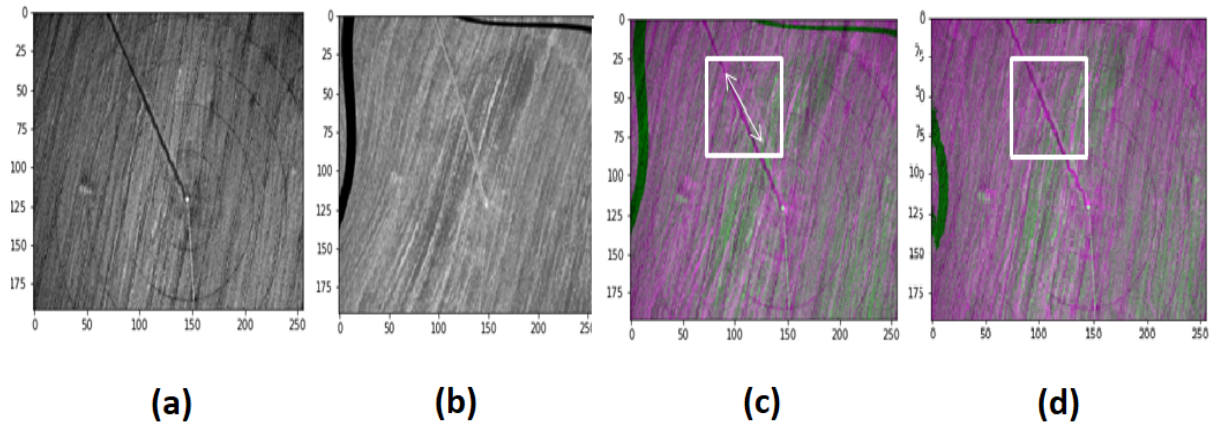


Figure 47 – Example of co-registration in cotton dataset: in (a) we have the fixed band (Green); in (b) we have the moving band (Red); in (c) we have the overlap of the bands (a) and (b); and in (d) the overlap of the fixed band with the moving band after the application of the field predicted by the network in (b).

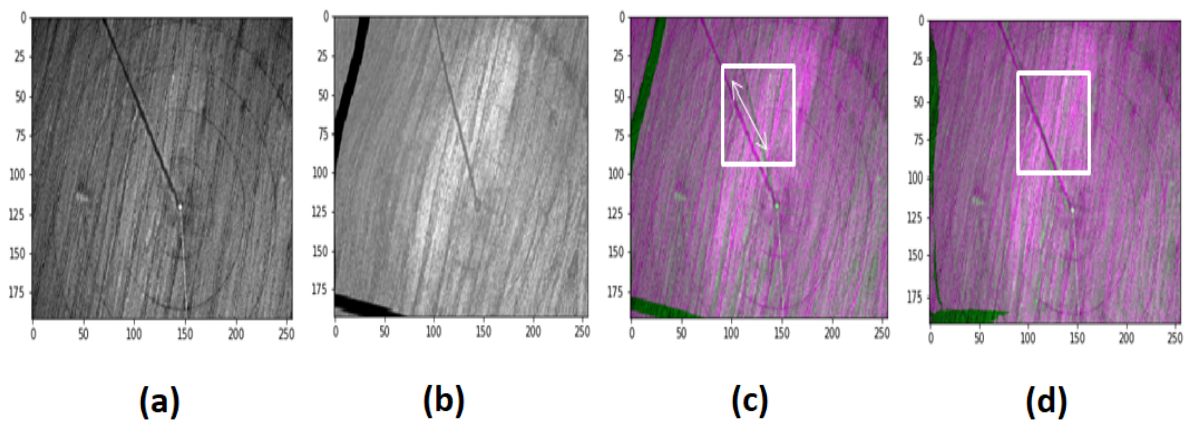


Figure 48 – Example of co-registration in cotton dataset: in (a) we have the fixed band (Green); in (b) we have the moving band (NIR); in (c) we have the overlap of the bands (a) and (b); and in (d) the overlap of the fixed band with the moving band after the application of the field predicted by the network in (b).

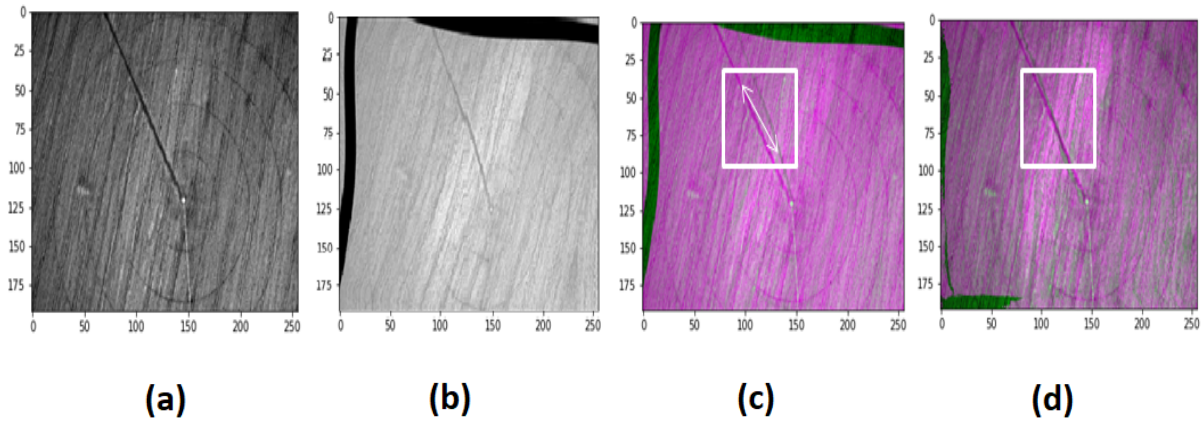


Figure 49 – Example of co-registration in cotton dataset: in (a) we have the fixed band (green); in (b) we have the moving band (Red Edge); in (c) we have the overlap of the bands (a) and (b); and in (d) the overlap of the fixed band with the moving band after the application of the field predicted by the network in (b).

Conclusion

This work presented the use of approaches based on deep learning for the identification and correction of deformations in images (including multispectral) obtained by UAV in PA.

Convolutional Neural Networks (proposed and traditional) proved to be efficient in the task of identifying the presence of linear and non-linear deformations in images obtained by UAV. Also, the proposed architecture, with less parameters and consequently lower computational cost, proved to be adherent in the task of identifying deformations. In this way, the proposed architecture can suggest the possibility of a system embedded in the UAV to identify the presence of deformations in each image capture.

We propose a framework based on deep learning to perform the registration between the bands of a multispectral image obtained by UAV. The experiments with this framework considered the presence of linear and non-linear deformations, which generates a complex scenario. However, despite the complexity, the proposed approach proved to be able to estimate the displacement field between the bands and thus perform the co-registration, without the need for manual marking of points.

6.1 Main contributions

The main contributions of this work are:

1. Demonstration of the capacity of CNN with great precision the presence or not of linear and non-linear deformations in images obtained by UAV.
2. Proposition of a framework based on deep learning for co-registration between bands of a multispectral image, which is capable of treating linear and non-linear deformations, without the need for manual point marking.

6.2 Limitations

As limitations, we present:

- ❑ This work considered for the identification and correction of deformations a random set of deformations and their combinations. These deformations may not represent a real scenario.
- ❑ Also, our analysis of co-registration was visual. Therefore, we chose to resize the images, preserving their proportions, however, this can invariably lead to the loss of information from the images.

6.3 Future Works

As future work, we intend to:

- ❑ Consider other deformations (and its variations) in images obtained by UAV.
- ❑ As the green band was always considered “fixed” in this work, it is intended to define a graph to perform the co-registration, so that the best matches are found between the bands.
- ❑ Evaluate the performance of the proposed framework with images obtained by other sensors than those considered in this work.
- ❑ Analyze quantitatively the co-registration results obtained by our proposal.

6.4 Contributions in Literature

From this work, two publications were produced in the literature:

- ❑ Silva LH, Júnior JD, Santos J, Mari JF, Escarpinati M, Backes A. Classification of uavs’ distorted images using convolutional neural networks. In: Anais do XVI Workshop de Visão Computacional. Porto Alegre, RS, Brasil: SBC, 2020. p. 98–103. DOI: 10.5753/wvc.2020.13488.
- ❑ Silva LH, Jr JD, Santos JF, Mari JF, Escarpinati MC, Backes AR. Non-linear Distortion Recognition in UAVs’ Images using Deep Learning. In Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP, ISBN 978-989-758-488-6, pages 447-454. DOI: 10.5220/0010309504470454

Two other works, not related to the main theme of this research, have also been published in the literature:

- Junior JD, Backes AR, Escarpinati MC, Silva LH, Costa BC, Avelar MH. UAV-Multispectral Sensed Data Band Co-Registration Framework. In 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Niteroi, Brazil, 2020, pp. 223-228. DOI: 10.1109/IWSSIP48289.2020.9145079.
- Junior JD, Backes AR, Escarpinati MC, Silva LH, Costa BC, Avelar MH. Assessing the Adequability of FFT-based Methods on Registration of UAV-Multispectral Images. In Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP, ISBN 978-989-758-402-2, pages 437-444. DOI: 10.5220/0009123704370444.

Another work, entitled *Deformable co-registration using deep learning*, is being prepared and will soon be submit for publication.

Bibliography

ABADI, M. et al. **TensorFlow: A system for large-scale machine learning**. 2016. Disponível em: <<https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi>>.

ALCANTARILLA, P. F.; BARTOLI, A.; DAVISON, A. J. Kaze features. In: FITZGIBBON, A. et al. (Ed.). **Computer Vision – ECCV 2012**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 214–227. ISBN 978-3-642-33783-3. Disponível em: <<https://doi.org/10.1007/978-3-642-33718-5>>.

ANDERSSON, O.; MARQUEZ, S. R. **A comparison of object detection algorithms using unmanipulated testing images: Comparing SIFT, KAZE, AKAZE and ORB**. Dissertação (Mestrado), 2016.

BAHETI, B. et al. Eff-unet: A novel architecture for semantic segmentation in unstructured environment. In: **Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops**. [s.n.], 2020. p. 358–359. Disponível em: <https://openaccess.thecvf.com/content_CVPRW_2020/html/w22/Baheti_Eff-UNet_A_Novel_Architecture_for_Semantic_Segmentation_in_Unstructured_Environment_CVPRW_2020_paper.html>.

BANERJEE, B. P.; RAVAL, S. A.; CULLEN, P. J. Alignment of uav-hyperspectral bands using keypoint descriptors in a spectrally complex environment. **Remote Sensing Letters**, Taylor Francis, v. 9, n. 6, p. 524–533, 2018. Disponível em: <<https://doi.org/10.1080/2150704X.2018.1446564>>.

BAY, H.; TUYTELAARS, T.; GOOL, L. V. Surf: Speeded up robust features. In: LEONARDIS, A.; BISCHOF, H.; PINZ, A. (Ed.). **Computer Vision – ECCV 2006**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006. p. 404–417. ISBN 978-3-540-33833-8. Disponível em: <https://link.springer.com/chapter/10.1007/11744023_32>.

BERNI, J. A. J. et al. Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. **IEEE Transactions on Geoscience and Remote Sensing**, v. 47, n. 3, p. 722–738, March 2009. ISSN 0196-2892. Disponível em: <<https://doi.org/10.1109/TGRS.2008.2010457>>.

BLACKMER, T. M.; SCHEPERS, J. S. Aerial photography to detect nitrogen stress in corn. **Journal of Plant Physiology**, Elsevier, v. 148, n. 3-4, p. 440–444, 1996. Disponível em: <[https://doi.org/10.1016/S0176-1617\(96\)80277-X](https://doi.org/10.1016/S0176-1617(96)80277-X)>.

CHRISTENSEN, G. E. et al. Individualizing neuro-anatomical atlases using a massively parallel computer. **Computer**, IEEE, v. 29, n. 1, p. 32–38, 1996. Disponível em: <<https://doi.org/10.1109/2.481434>>.

DALL’AGNOL, A.; MOREIRA, P. **Parece mentira, mas o agro da Holanda supera o do Brasil em exportação**. 2019. Disponível em: <<https://blogs.canalrural.com.br/embrapasoja/2019/03/06/parece-mentira-mas-o-agronegocio-da-holanda-supera-o-do-brasil/>>. Acesso em: 05/02/2021.

Dias Junior, J. D. et al. Uav-multispectral sensed data band co-registration framework. In: **2020 International Conference on Systems, Signals and Image Processing (IWSSIP)**. [s.n.], 2020. p. 223–228. Disponível em: <<https://doi.org/10.1109/IWSSIP48289.2020.9145079>>.

DYRMANN, M.; KARSTOFT, H.; MIDTIBY, H. S. Plant species classification using deep convolutional neural network. **Biosystems Engineering**, Elsevier, v. 151, p. 72–80, 2016. Disponível em: <<https://doi.org/10.1016/j.biosystemseng.2016.08.024>>.

EPPENHOF, K. A. et al. Progressively trained convolutional neural networks for deformable image registration. **IEEE Transactions on Medical Imaging**, IEEE, v. 39, n. 5, p. 1594–1604, 2019. Disponível em: <<https://doi.org/10.1109/TMI.2019.2953788>>.

EPPENHOF, K. A. J.; PLUIM, J. P. W. Pulmonary ct registration through supervised learning with convolutional neural networks. **IEEE Transactions on Medical Imaging**, v. 38, n. 5, p. 1097–1105, 2019. Disponível em: <<https://doi.org/10.1109/TMI.2018.2878316>>.

FARMER, B. Perspectives on the ‘green revolution’ in south asia. **Modern Asian Studies**, Cambridge University Press, v. 20, n. 1, p. 175–199, 1986.

FAUSETT, L. V. **Fundamentals of neural networks: architectures, algorithms and applications**. [S.l.]: Pearson Education India, 2006.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing (3rd Edition)**. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X.

GOODFELLOW, I. et al. **Deep learning**. [S.l.]: MIT press Cambridge, 2016. v. 1.

GUO, Y. et al. Deep learning for visual understanding: A review. **Neurocomputing**, v. 187, p. 27–48, 2016. Disponível em: <https://doi.org/10.1111/clr.46_12957>.

HAN, Y. et al. Automated coregistration of multisensor orthophotos generated from unmanned aerial vehicle platforms. **Journal of Sensors**, Hindawi, v. 2019, 2019. Disponível em: <<https://doi.org/10.1155/2019/2962734>>.

HASSOUN, M. H. et al. **Fundamentals of artificial neural networks**. [S.l.]: MIT press, 1995.

HAYKIN, S.; NETWORK, N. A comprehensive foundation. **Neural networks**, v. 2, n. 2004, p. 41, 2004. Disponível em: <<https://doi.org/10.1002/wilm.42820040110>>.

HAZELL, P. B. **The Asian green revolution**. [S.l.]: Intl Food Policy Res Inst, 2009. v. 911.

He, K. et al. Deep residual learning for image recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [s.n.], 2016. p. 770–778. Disponível em: <<https://doi.org/10.1109/CVPR.2016.90>>.

HEO, S.-J. et al. Deep learning algorithms with demographic information help to detect tuberculosis in chest radiographs in annual workers' health examination data. **International journal of environmental research and public health**, Multidisciplinary Digital Publishing Institute, v. 16, n. 2, p. 250, 2019. Disponível em: <<https://doi.org/10.3390/ijerph16020250>>.

HUNT, E. R. et al. Acquisition of nir-green-blue digital photographs from unmanned aircraft for crop monitoring. **Remote Sensing**, v. 2, p. 290–305, 2010. Disponível em: <<https://doi.org/10.3390/rs2010290>>.

HUNTER, M. C. et al. Agriculture in 2050: Recalibrating Targets for Sustainable Intensification. **BioScience**, v. 67, n. 4, p. 386–391, 02 2017. ISSN 0006-3568. Disponível em: <<https://doi.org/10.1093/biosci/bix010>>.

IANDOLA, F. N. et al. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size. **arXiv:1602.07360**, 2016. Disponível em: <<https://arxiv.org/abs/1602.07360>>.

IBGE. **Levantamento Sistemático da Produção Agrícola - LSPA**. 2020. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/9201-levantamento-sistematico-da-producao-agricola.html>>. Acesso em: 12/09/2020.

_____. **Produto Interno Bruto - PIB**. 2020. Disponível em: <<https://www.ibge.gov.br/explica/pib.php>>. Acesso em: 12/09/2020.

IGLOVIKOV, V.; SHVETS, A. Terausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. **arXiv preprint arXiv:1801.05746**, 2018. Disponível em: <<https://arxiv.org/abs/1801.05746>>.

IHUOMA, S. O.; MADRAMOOTOO, C. A. Recent advances in crop water stress detection. **Computers and Electronics in Agriculture**, Elsevier, v. 141, p. 267–275, 2017. Disponível em: <<https://doi.org/10.1016/j.compag.2017.07.026>>.

JENKINS, D.; VASIGH, B. **The economic impact of unmanned aircraft systems integration in the United States**. [S.l.]: Association for Unmanned Vehicle Systems International (AUVSI), 2013.

JUNIOR, J. D. D.; BACKES, A. R.; ESCARPINATI, M. C. Detection of control points for uav-multispectral sensed data registration through the combining of feature descriptors. In: **VISIGRAPP (4: VISAPP)**. [s.n.], 2019. p. 444–451. Disponível em: <<https://www.scitepress.org/Papers/2019/75802/75802.pdf>>.

JUNIOR, J. D. D. et al. Uav-multispectral sensed data band co-registration framework. Universidade Federal de Uberlândia, 2020.

KATAOKA, T. et al. Crop growth estimation system using machine vision. In: **IEEE Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)**. 2003. v. 2, p. b1079–b1083. Disponível em: <<https://doi.org/10.1109/AIM.2003.1225492>>.

KHANAL, S.; FULTON, J.; SHEARER, S. An overview of current and potential applications of thermal remote sensing in precision agriculture. **Computers and Electronics in Agriculture**, Elsevier, v. 139, p. 22–32, 2017. Disponível em: <<https://doi.org/10.1016/j.compag.2017.05.001>>.

LECUN, Y.; BENGIO, Y.; HINTON, G. E. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. Disponível em: <<https://doi.org/10.1038/nature14539>>.

LECUN, Y. et al. Handwritten digit recognition with a back-propagation network. In: **Advances in neural information processing systems**. [s.n.], 1990. p. 396–404. Disponível em: <<https://proceedings.neurips.cc/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf>>.

LECUN, Y. L. et al. Gradient-based learning applied to document recognition. **Proceedings of IEEE**, v. 86, n. 11, p. 2278–2324, nov. 1998. Disponível em: <<https://doi.org/10.1109/5.726791>>.

LEUTENEGGER, S.; CHLI, M.; SIEGWART, R. Y. Brisk: Binary robust invariant scalable keypoints. In: **2011 International Conference on Computer Vision**. [s.n.], 2011. p. 2548–2555. ISSN 2380-7504. Disponível em: <<https://doi.org/10.1109/ICCV.2011.6126542>>.

LONG, J.; SHELHAMER, E.; DARRELL, T. Fully convolutional networks for semantic segmentation. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [s.n.], 2015. p. 3431–3440. Disponível em: <https://openaccess.thecvf.com/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html>.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, v. 60, n. 2, p. 91–110, Nov 2004. ISSN 1573-1405. Disponível em: <<https://doi.org/10.1023/B:VISI.0000029664.99615.94>>.

MAES, W. H.; STEPPE, K. Perspectives for remote sensing with unmanned aerial vehicles in precision agriculture. **Trends in plant science**, Elsevier, v. 24, n. 2, p. 152–164, 2019. Disponível em: <<https://doi.org/10.1016/j.tplants.2018.11.007>>.

MALTHUS, T. R. **An Essay on the Principle of Population..** [S.l.: s.n.], 1872.

MALTHUS, T. R.; WINCH, D.; JAMES, P. **Malthus: 'An Essay on the Principle of Population'**. [S.l.]: Cambridge University Press, 1992.

MARCOS, A. P.; RODOVALHO, N. L. S.; BACKES, A. R. Coffee leaf rust detection using convolutional neural network. In: IEEE. **2019 XV Workshop de Visão Computacional (WVC)**. 2019. p. 38–42. Disponível em: <<https://doi.org/10.1109/WVC.2019.8876931>>.

_____. Coffee leaf rust detection using genetic algorithm. In: IEEE. **2019 XV Workshop de Visão Computacional (WVC)**. 2019. p. 16–20. Disponível em: <<https://doi.org/10.1109/WVC.2019.8876934>>.

MATHWORKS. **Align multiple scenes into a single image using image registration**. 2018. <https://www.mathworks.com/discovery/image-registration.html> [Online; retirado em 15 de fevereiro de 2020 às 17:18].

- MCBRATNEY, A. et al. Future directions of precision agriculture. **Precision agriculture**, Springer, v. 6, n. 1, p. 7–23, 2005. Disponível em: <<https://doi.org/10.1007/s11119-005-0681-8>>.
- MILELLA, A.; REINA, G.; NIELSEN, M. A multi-sensor robotic platform for ground mapping and estimation beyond the visible spectrum. **Precision agriculture**, Springer, v. 20, n. 2, p. 423–444, 2019. Disponível em: <<https://doi.org/10.1007/s11119-018-9605-2>>.
- MILICS, G. Application of uavs in precision agriculture. In: **International Climate Protection**. Springer, 2019. p. 93–97. Disponível em: <https://doi.org/10.1007/978-3-030-03816-8_13>.
- PENA, R. F. A. **Explosão Demográfica**. 2013. Disponível em: <<https://brasilecola.uol.com.br/geografia/explosao-demografica.htm>>. Acesso em: 29/10/2020.
- PONTI, M. A. et al. Everything you wanted to know about deep learning for computer vision but were afraid to ask. In: **SIBGRAPI Tutorials**. IEEE Computer Society, 2017. p. 17–41. ISBN 978-1-5386-0619-3. Disponível em: <<https://doi.org/10.1109/SIBGRAPI-T.2017.12>>.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **International Conference on Medical image computing and computer-assisted intervention**. 2015. p. 234–241. Disponível em: <https://doi.org/10.1007/978-3-319-24574-4_28>.
- SANKARAN, S. et al. Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping: A review. **European Journal of Agronomy**, Elsevier, v. 70, p. 112–123, 2015. Disponível em: <<https://doi.org/10.1016/j.eja.2015.07.004>>.
- SHEKHAR, R.; ZAGRODSKY, V. Mutual information-based rigid and nonrigid registration of ultrasound volumes. **IEEE transactions on medical imaging**, IEEE, v. 21, n. 1, p. 9–22, 2002. Disponível em: <<https://doi.org/10.1109/42.981230>>.
- SILVA, L. et al. Non-linear distortion recognition in uavs' images using deep learning. In: INSTICC. **Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP**. SciTePress, 2021. p. 447–454. ISBN 978-989-758-488-6. Disponível em: <<https://doi.org/10.5220/0010309504470454>>.
- _____. Classification of uavs' distorted images using convolutional neural networks. In: **Anais do XVI Workshop de Visão Computacional**. Porto Alegre, RS, Brasil: SBC, 2020. p. 98–103. Disponível em: <<https://doi.org/10.5753/wvc.2020.13488>>.
- SILVA, R. R. d. et al. Detection of sugarcane crop rows from uav images using semantic segmentation and radon transform. Universidade Federal de Uberlândia, 2020.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014. Disponível em: <<https://arxiv.org/abs/1409.1556>>.
- Szegedy, C. et al. Rethinking the inception architecture for computer vision. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [s.n.], 2016. p. 2818–2826. Disponível em: <<https://doi.org/10.1109/CVPR.2016.308>>.

UCHIDA, S. Image processing and recognition for biological images. In: **Development, growth & differentiation**. [s.n.], 2013. Disponível em: <<https://doi.org/10.1111/dgd.12054>>.

UN, U. N. **COVID-19 Response**. 2020. Disponível em: <<https://www.un.org/en/coronavirus>>. Acesso em: 20/10/2020.

_____. **Population Division**. 2020. Disponível em: <<https://www.un.org/development/desa/pd/>>. Acesso em: 20/10/2020.

VOS, B. D. de et al. A deep learning framework for unsupervised affine and deformable image registration. **Medical image analysis**, Elsevier, v. 52, p. 128–143, 2019. Disponível em: <<https://doi.org/10.1016/j.media.2018.11.010>>.

WALIMBE, V.; SHEKHAR, R. Automatic elastic image registration by interpolation of 3d rotations and translations from discrete rigid-body transformations. **Medical Image Analysis**, Elsevier, v. 10, n. 6, p. 899–914, 2006. Disponível em: <<https://doi.org/10.1016/j.media.2006.09.002>>.

WANG, Y.; STAIB, L. H. Elastic model based non-rigid registration incorporating statistical shape information. In: SPRINGER. **International Conference on Medical Image Computing and Computer-Assisted Intervention**. 1998. p. 1162–1173. Disponível em: <<https://doi.org/10.1007/BFb0056306>>.

YASIR, R. et al. Data-driven multispectral image registration. In: IEEE. **2018 15th Conference on Computer and Robot Vision (CRV)**. 2018. p. 230–237. Disponível em: <<https://doi.org/10.1109/CRV.2018.00040>>.

ZARCO-TEJADA, P. J. et al. New tools and methods in agronomy. In: **Principles of Agronomy for Sustainable Agriculture**. Springer, 2016. p. 503–514. Disponível em: <https://doi.org/10.1007/978-3-319-46116-8_33>.