ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

SELF-SUPERVISED PANSHARPENING: GUIDED COLORIZATION OF PANCHROMATIC IMAGES USING GENERATIVE ADVERSARIAL NETWORKS

M.Sc. THESIS

Furkan ÖZÇELİK

Department of Computer Engineering

Computer Engineering Programme

JULY 2020

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

SELF-SUPERVISED PANSHARPENING: GUIDED COLORIZATION OF PANCHROMATIC IMAGES USING GENERATIVE ADVERSARIAL NETWORKS

M.Sc. THESIS

Furkan ÖZÇELİK
(504181516)

Department of Computer Engineering

Computer Engineering Programme

Thesis Advisor: Prof. Dr. Gözde ÜNAL

JULY 2020

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

ÖZ-DENETİMLİ PANKESKİNLEŞTİRME: ÇEKİŞMELİ ÜRETKEN AĞLAR İLE PANKROMATİK GÖRÜNTÜLERİN GÜDÜMLÜ RENKLENDİRİLMESİ

YÜKSEK LİSANS TEZİ

Furkan ÖZÇELİK
(504181516)

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Gözde ÜNAL

TEMMUZ 2020

Furkan ÖZÇELİK, a M.Sc. student of İTU Graduate School of Science Engineering and Technology student ID 504181516, successfully defended the thesis entitled "SELF-SUPERVISED PANSHARPENING: GUIDED COLORIZATION OF PANCHROMATIC IMAGES USING GENERATIVE ADVERSARIAL NETWORKS", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Gözde ÜNAL** İstanbul Technical University	....
Jury Members :	**Prof.Dr. Hazım Kemal EKENEL** Istanbul Technical University	....
	Dr.Öğr.Üye. Ramazan Gökberk CİNBİŞ Middle East Technical University	....

Date of Submission : 15 June 2020
Date of Defense : 09 July 2020

To my family and friends,

FOREWORD

This work was supported by Research Fund of the Istanbul Technical University Project Number: MGA-2017-40811. This thesis was supported by the Turkcell-Istanbul Technical University Researcher Funding Program.

I would like to thank Turkcell-ITU Scholarship Programme for their generous funding during my master’s study, and professors Uğur Algancı and Elif Sertel at ITU-CSCRS for providing support for this research.

I would like to thank my advisor Prof. Dr. Gözde Ünal for her amazing support and guidance throughout this research, and inspiration to love academic research. I would like to express my sincere gratitude to my family and friends who stood by me through thick and thin.

July 2020 Furkan ÖZÇELİK

TABLE OF CONTENTS

Page

ABBREVIATIONS

GAN	: Generative Adversarial Networks
RaGAN	: Realistic Average GAN
ANN	: Artificial Neural Network
CNN	: Convolutional Neural Network
PAN	: Panchromatic Image
MS	: Multispectral Image
PS	: Pansharpened Image
GMS	: Grayscale Multispectral Image
SR	: Super-Resolution
DL	: Deep Learning
ReLU	: Rectified Linear Unit
CS	: Component Substitution
MRA	: Multiresolution Analysis
ATWT	: A trous Wavelet Transform
BDSD	: Band-Dependent Spatial Detail Algorithm
GSA	: Adaptive Gram Schmidt Algorithm
IHS	: Intensity Hue Saturation Method
PCA	: Principal Component Analysis
TA-CNN	: Target Adaptive CNN
PSNR	: Peak Signal-to-Noise Ratio
sCC	: Spatial Correlation Coefficient
SSIM	: Structural Similarity Index
Q	: Quality Index
SAM	: Spectral Angle Mapper
ERGAS	: Relative Dimensionless Global Error in Synthesis
QNR	: Quality with No Reference

SYMBOLS

G	: Generator
D	: Discriminator
X	: Input
X_{MS}	: Reduced-Resolution Multispectral Input Image
X_{GMS}	: Grayscale Multispectral Input Image
X_{PAN}	: Reduced-Resolution Panchromatic Input Image
Y	: Ground Truth
Y_{MS}	: Original-Resolution Multispectral Image
Y_{PAN}	: Original-Resolution Panchromatic Image
$\hat{Y}$	: Predicted Value for Pansharpened Image
$\hat{Y}_P$	: Pansharpened Image (with Reduced-Resolution PAN)
$\hat{Y}_G$	: Colorized Image (GMS colorized with Reduced-Resolution MS)
$\hat{Y}_F$	: Full-Resolution Pansharpened Image
$\sigma(.)$	: Sigmoid Function
$\downarrow 4\times 4$	: $4\times$ Downsampling
$\uparrow 4\times 4$	: $4\times$ Upsampling

LIST OF TABLES

Page

LIST OF FIGURES

Page

**SELF-SUPERVISED PANSHARPENING:
GUIDED COLORIZATION OF PANCHROMATIC IMAGES
USING GENERATIVE ADVERSARIAL NETWORKS**

SUMMARY

Satellite images provide images with different properties. Multispectral images have low spatial resolution and high spectral resolution. Panchromatic images have high spatial resolution and low spectral resolution. The fusion process of these two images is called pansharpening. For decades, traditional image processing methods are designed for this process. After the inspirational success of Convolutional Neural Networks(CNN) in computer vision, CNN models are also designed for pansharpening.

Convolutional Neural Networks (CNN)-based approaches have shown promising results in pansharpening of satellite images in recent years. However, they still exhibit limitations in producing high-quality pansharpening outputs. We identified a spatial detail disagreement problem between reduced resolution panchromatic images and original multispectral images, which are assumed to have the same resolution. This problem causes an insufficient training process in current CNN-based pansharpening models.

We propose a new self-supervised learning framework, where we treat pansharpening as a colorization problem, which brings an entirely novel perspective and solution to the problem compared to existing methods that base their solution solely on producing a super-resolution version of the multispectral image. CNN-based methods provide a reduced resolution panchromatic image as input to their model along with reduced resolution multispectral images, hence learn to increase their resolution together. In the training phase of our model, reduced resolution panchromatic image is substituted with grayscale transformed multispectral image, thus our model learns colorization of the grayscale input.

We further address the fixed downscale ratio assumption during training, which does not generalize well to the full-resolution scenario. We introduce a noise injection into the training by randomly varying the downsampling ratios. Those two critical changes, along with the addition of adversarial training in the proposed PanColorization Generative Adversarial Networks (PanColorGAN) framework, help overcome the spatial detail loss and blur problems that are observed in CNN-based pansharpening. The proposed approach outperforms the previous CNN-based and traditional methods as demonstrated in our experiments.

ÖZ-DENETİMLİ PANKESKİNLEŞTİRME: ÇEKİŞMELİ ÜRETKEN AĞLAR İLE PANKROMATİK GÖRÜNTÜLERİN GÜDÜMLÜ RENKLENDİRİLMESİ

ÖZET

Uzaktan algılama amaçlı uydu sistemlerinde hem uzamsal hem de spektral çözünürlüğü yüksek görüntüler üretmek önemli bir görevdir. Bu işlem için tek bir sensör yeterli olmayacağı için Pleiades, GeoEye, Quickbird ve Worldview gibi birçok uyduda pankromatik ve multispektral görüntü elde eden sensörler bulunur. Pankromatik sensörler uzamsal çözünürlüğe odaklanıp tek kanallı bir yapıya sahip görüntü oluştururken, multispektral sensörler spektral çözünürlüğe odaklanıp çok kanallı bir yapıya sahip görüntü oluştururlar. Pankromatik sensörlerin elde ettiği görüntülerin uzamsal çözünürlükleri genelde aynı uyduda bulunan multispektral sensörden 4 kat daha fazladır. Pankromatik ve multispektral görüntülerin belirli algoritmalar yoluyla füzyonuna da pankeskinleştirme (pansharpening) denmektedir.

Henüz Evrişimsel Sinir Ağları (Convolutional Neural Networks) bu işlem için kullanılmadan önce birçok farklı algoritma önerilmiştir. Bunlara geleneksel pankeskinleştirme metotları diyoruz. Bu metotlar Bileşen Değiştirmeli (Component Substitution) metotlar ve Çoklu Çözünürlük Analizi (Multiresolution Analysis) metotları olmak üzere ikiye ayrılıyor. Bileşen değiştirmeli metotlar multispektral resmi uzamsal ve spektral bileşenlerine ayırıp uzamsal bileşeni pankromatik resimden elde edilen bir uzamsal bileşenle değiştirme üzerine bina edilmiştir. Bu metotların arasında Yeğinlik-Renk-Doyum (IHS), Temel Bileşen Analizi (PCA), Gram-Schmidt, Brovey Dönüşümü, BDSD ve PRACS gibi birçok metot bulunmaktadır. Çoklu çözünürlük analizi metotları ise pankromatik resimler üzerinden sinyal filtreleri geçirilerek elde edilen bilgilerin multispektal resimlere enjekte edilmesiyle oluşuyor. Bu metotlara örnek olarak da yüksek-geçiren filtreler (HPF), modülasyon transfer fonksiyonu (MTF) bazlı Genelleştirilmiş Laplasyen piramitleri (MTF-GLP), Yüksek geçiren modülasyonlu MTF-GLP (MTF-GLP-HPM), uzamsal temel bileşen ayrıştırma (SPCA) ve dalgacık dönüşümlü (wavelet transform) metotlar sayılabilir.

Son yıllarda evrişimsel sinir ağlarının başarılarıyla derin öğrenme metotları birçok alanda uygulanmaya başlamıştır. Bu gelişimi sağlayan en büyük unsurlar donanımda yaşanan geliştirmeler, verinin artması ve sinir ağları mimarilerinde çeşitli problemlerin çözülmesi olmuştur. Bu ilerlemelerle birlikte bilgisayar görüsü, ses tanıma, doğal dil işleme, medikal görünteleme ve robotik gibi birçok alanda derin öğrenme modelleri kullanılmaya başlanmıştır. Derin öğrenmenin sınıflandırmaya dair başarılarıyla birlikte sentezlemeye dair de birçok başarısı olmuştur. Özellikle görüntü sentezlemede evrişimsel sinir ağları yapısına sahip Çekişmeli Üretken Ağlar (Generative Adversarial Networks) gerçekçi üretim kapasitesiyle öne çıkmış ve birçok alanda kullanılmaya başlamıştır.

Çekişmeli üretken ağlar (ÇÜA), belirli bir hedef dağılıma sahip resimlerin sentezlenilmesini öğrenebilen üretken modellerin bir sınıfını oluşturur. Genelde ÇÜA

modelinde iki yapay sinir ağı modeli bulunur. Bunlardan biri üretici ağ diğeri ise ayrıştırıcı ağdır. Resim uygulaması üzerinden konuşulacak olursa, üretici ağ rastgele örneklenmiş sayıları girdi olarak alır ve bunları ağ içinde işleyerek bir resim üretir. Ayrıştırıcı ağ ise hem üretken ağın ürettiği resimlerden hem de elde bulunan veri setinden belirli sayıda resim alır ve bunları gerçek ve sahte olarak sınıflandırmaya çalışır. Ayrıştırıcı ağın yaptığı doğru ve yanlış sınıflandırmalara göre hem ayrıştırıcı ağ hem de üretici ağ eğitilir. Ayrıştırıcı ağ veri kümesinden gelen resimleri gerçek, üretici ağın ürettiklerini sahte olarak sınıflandırmaya çalışırken, üretici ağ ise kendi ürettiği resimlerin ayrıştırıcı ağ tarafından gerçek olarak sınıflandırılması için çabalar. ÇÜA modeli geliştirilerek oluşturulmuş Pix2Pix modeli görüntüden görüntüye dönüşüm (image-to-image translation) uygulamalarında büyük başarı göstermiştir.

Evrişimli sinir ağları ve çekişmeli üretken ağların görüntü sentezlemedeki bu başarısı pankeskinleştirme çalışmalarında da kendini göstermiştir. Şu ana kadar yapılan evrişimli sinir ağları tabanlı pankeskinleştirme çalışmalarında görüntüden görüntüye dönüşüm uygulamalarından biri olan çözünürlük arttırmadan (super-resolution) esinlenilmiştir. Tam-çözünürlükte referans resmi olmadığı için, bu modellerin eğitimi sırasında çözünürlüğü düşürülmüş pankromatik görüntü ve çözünürlüğü düşürülmüş multispektral görüntü girdi olarak verilerek, çıktı olarak normal multispektal görüntü elde edilmeye çalışılmaktadır. Çözünürlük arttırmadan esinlenerek tasarlanılan bu modellerin çeşitli yetersizlikleri bulunmaktadır. Analizlerimiz sonucunda bu tarz bir yaklaşımın pankromatik görüntülerin uzamsal çözünürlüklerini yeterince iyi aktaramadığını ve tam-çözünürlük üzerinde denendiğinde bulanıklık problemleri olduğunu tespit ettik.

Çözünürlük arttırmadan esinlenmiş modellerin bu problemlerini tespit etmekle birlikte başka bir görüntüden görüntüye dönüşüm uygulaması olan renklendirmeden (colorization) esinlenen bir model olarak PanColorGAN modelini tasarladık. Çözünürlüğü düşürülmüş multispektral görüntünün yanında girdi olarak çözünürlüğü düşürülmüş pankromatik resmi vermek yerine normal çözünürlükteki multispektral resmin renksiz halini verdik. Çıktı olarak da normal multispektral resmi beklediğimizden, girdi ve çıktı arasındaki uzamsal detay farklılığı problemini çözmüş olduk. Bununla birlikte hala tam-çözünürlük senaryosundaki bulanıklık problemi tamamiyle çözülmediği için rastgele altörnekleme (random downsampling) dediğimiz bir yöntemi de modelimize ekledik. Bu yöntemle eğitim sırasında multispektral görüntünün çözünürlüğünü düşürürken belirli bir sayıya değil değil de belirli bir aralıktan rastgele seçilmiş bir sayıya düşürüyoruz. Bu adeta bir gürültü enjeksiyonu görevi gördüğü için modelimizin istemediğimiz bir görevi ezberlemediğinden ve gerçekten renklendirmeyi öğrendiğinden emin olmamızı sağlıyor. Bu sayede farklı çözünürlüklere, koşullara ve uydu yapılarına karşı gürbüz bir yapıya sahip olan bir model geliştirmiş olduk.

Eğitimlerimizi Uydu Haberleşme ve Uzaktan Algılama Merkezi'nin (UHUZAM) bizlere sağladığı Pleiades uydusundan elde edilen Türkiye'nin farklı şehirleri üzerinde yaptık. Aydın, İstanbul, Bursa, Bilecik, Muğla illerinden elde edilmiş bu görüntüler 2 metre multispektral ve 0.5 metre pankromatik çözünürlüklerine sahiptir. Bu görüntülerin oluşturduğu büyük çerçeveleri sinir ağlarına uygun hale getirebilmek için birbirine denk gelen 1024x1024'lük pankromatik resimlerine ve 256x256'lık multispektral resimlerine ayırdık. Bu ayrıştırma sonucu Pleiades'ten 30000 eğitim örneği ve 5700 test örneği elde ettik. Ayrıca Digital Globe firmasının da, görüntülerini Worldview 2 ve Worldview 3 uydularından elde ederek sunduğu ücretsiz bir veriseti üzerinde de modelimizin farklı koşullarda nasıl çalıştığını test ettik. Pleiades ile benzer bir ayrıştırma işlemi uyguladığımız bu verisetinden de sadece test için kullanılmak üzere 350 görüntü elde ettik.

Tasarladığımız PanColorGAN modelini rastgele altörnekleme kullanan ve kullanmayan iki versiyonla test ettik, altörnekleme kullanan versiyona PanColorGAN+RD adını verdik. Ayrıca PanColorGAN ile aynı mimari yapıya sahip olup çözünürlük arttırmadan esinlenilmiş metoda sahip olan PanSRGAN modelini de karşılaştırma amaçlı test ettik. Bu modellerin yanında daha ayrıntılı bir değerlendirme için geleneksel pankeskinleştirme modellerinden olan BDSD, ATWT, GSA, GLP-REG-FS, Nonlinear IHS, Semi-blind Convolution gibi modelleri ve ayrıca çözünürlük arttırmadan esinlenilmiş evrişimli sinir ağları modelleriyle oluşturulmuş olan PanNet ve TA-CNN modellerini de aynı verisetleri üzerinde test ettik.

Sayısal değerlendirmelerde düşürülmüş-çözünürlükte referanslı metrikler olan QAVE, SAM, ERGAS, sCC ve Q metriklerini, tam-çözünürlükte ise referanssız metrikler olarn D_s, D_λ ve QNR metriklerini kullandık. Birçok model için sayısal verilerin zaman zaman görsel sonuçlarla uyuşmadığını gördük. Bunun görüntü işlemede genel bir durum olduğunu biliyoruz ki literatürü incelediğimizde pankeskinleştirme için de durumun böyle olduğunu anladık. İki uydunun hem düşürülmüş-çözünürlük hem de tam-çözünürlük koşullarında da görsel olarak PanColorGAN modelimiz en iyi sonuçları elde etmeyi başardı ve genelleştirilebilirlik yönüyle en iyi performansı da gösterdi.

1. INTRODUCTION

In this thesis, we present a novel deep learning-based model for the task known as pansharpening[1].

1.1 Pansharpening

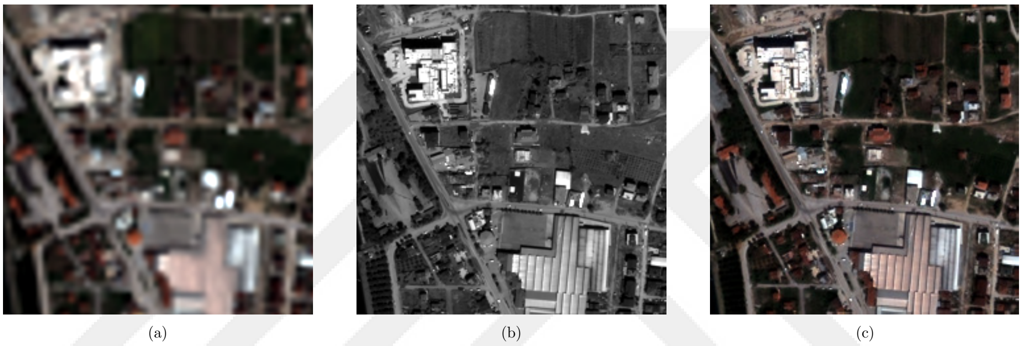

Figure 1.1 : Example to input and output images of pansharpening: (a) Multispectral Image (b) Panchromatic Image (c) Pansharpened Image.

Designing algorithms to obtain images with high-quality properties both in spatial and spectral domains is an significant task in remote sensing.

In remote sensing, it is a significant task to design algorithms to obtain images with high-quality properties both spatially and spectrally. As a single sensor is not sufficient to get dual-domain high-resolution images, many of the satellites such as Pleiades, GeoEye, Quickbird, and Worldview constellations contain both panchromatic and multispectral sensors. Panchromatic sensors focus on spatial resolution while providing images with a single-band, whereas multispectral sensors focus on spectral resolution while providing multi-band images. The fusion of these two modalities with a prescribed algorithm in order to obtain high-resolution images

[1] ©2020 IEEE. Text excerpts, figures and tables in this thesis are reproduced, with permission, from Ozcelik, F., Alganci, U., Sertel, E. and Unal, G. 2020. Rethinking CNN-Based Pansharpening: Guided Colorization of Panchromatic Images Via GANs, *IEEE Transactions on Geoscience and Remote Sensing* (Accepted)

in both domains is known as pansharpening. Figure 1.1 depicts the pansharpening problem.

Many different approaches are designed for pansharpening task before deep learning-based methods are employed. We refer to them as traditional pansharpening methods. Traditional pansharpening algorithms mainly consist of two categories which are component substitution based methods and multiresolution analysis methods. We further describe these methods in Chapter 2.

In this thesis, we propose a novel deep learning-based pansharpening method (PanColorGAN), which significantly advances the state-of-the-art in pansharpening. We introduce deep learning and its place in remote sensing in this chapter, and detail those later in the Background Chapter 2. Furthermore, we provide an extended motivation to our proposal in Chapter 3. The methodological details of the PanColorGAN framework are provided in Chapter 4.

1.2 Deep Learning

Recent availability of large datasets, increased computing power, advanced architectures and optimization led the way to the adaptation of deep learning techniques to numerous problems in computer vision as well as in remote sensing. Typically, a dedicated convolutional neural network (CNN) model is built in order to learn specific supervised learning tasks such as classification and detection, and lately to learn unsupervised learning tasks, particularly in image generation problems. For the latter, generative models such as Convolutional Autoencoders and Generative Adversarial Networks (GANs) [2] are applied to self-supervised image synthesis tasks such as Super-Resolution (SR) [3] and Colorization [4]. The self-supervision in SR models is realized by reducing the resolution of the input ($2\times$ - $4\times$ times typically) during training and allowing the network model to learn to increase the resolution of the input images. The reconstruction loss between the output of the network and the original image is calculated, which is used in the optimization of the network parameters. Colorization is another popular self-supervised synthesis task encountered in computer vision. This time, the network tries to learn to colorize grayscale images, which are created from their color counterparts in the training phase. As the network tries to reconstruct original color images, the corresponding loss between the output of the network and the original image is utilized in the network optimization process.

1.3 Deep Learning in Remote Sensing

In the field of remote sensing, in addition to widely-studied supervised learning problems such as land cover classification, building detection, deep network models are recently applied to the pansharpening task [5]. Existing CNN-based pansharpening methods in the literature [6–14] can be re-interpreted in the framework of self-supervised learning for the super-resolution task while following the commonly used Wald's protocol [15]. Although they differ in many aspects, all CNN-based methods have some common properties in the training procedure. In the training phase, pansharpening models are provided with the reduced resolution panchromatic and reduced resolution multispectral image as inputs in order to learn to reconstruct a high-resolution multispectral image at the output. Inspired by Wald's protocol, all previous studies treated CNN-based pansharpening only as a super-resolution task. However, we hypothesize and show in this thesis that using another self-supervised learning task, namely colorization, is more suitable to the pansharpening problem.

1.4 Our Method and Contributions

The motivation behind our introducing a colorization-based self-supervised learning approach to pansharpening is based on our observations of an inefficient level of spatial-detail-preservation in the former approaches. We demonstrate this problem and describe why it is encountered in Chapter 3. As a solution, we present a novel pansharpening approach, along with a new GAN-based dedicated colorization model, which we call PanColorGAN in Chapter 4. In Chapter 5, we present the results of the new method, which demonstrates an improved quantitative and qualitative performance, along with discussions, followed by conclusions in Chapter 6.

The contribution of our work are as follows:

- Spatial-detail differences between reduced resolution panchromatic images and full resolution multispectral images are presented.

- Effects of the spatial-detail difference on former CNN-based methods (inspired by super-resolution) are analyzed.

- The blurring problem which is caused by a "fixed upsample scale", is identified in full-resolution pansharpening.

- We propose a pansharpening framework inspired by the colorization self-supervision task.

- We design a GAN-based model called PanColorGAN which utilizes the pansharpening framework that is inspired by colorization. We also establish a color injection head in our network.

- We introduce the "Random Downsampling" method for solving the blurring problem that is caused by the "fixed upsample scale" assumption.

- Our proposed model is compared to several state-of-the-art methods, in order to demonstrate its performance both quantitatively and qualitatively.

2. BACKGROUND AND LITERATURE REVIEW

In this chapter, we give a prerequisite knowledge to understand the thesis. First, we describe relevant deep learning models. Secondly, we illustrate two self-supervision task in deep learning with images, which are image super-resolution and image colorization. Lastly, We represent pansharpening models that are devised before our model.

2.1 Deep Learning Models

After AlexNet [16] won the ILSVRC competition [17] in 2012, the ubiquity of Deep Learning methods has grown rapidly and applied to many different domains. The advantage of Deep Learning models compared to older Machine Learning models is that they do not require any hand-crafted features. With more data and computation provided, deep learning models can learn features at low, medium, and high levels. Different models are proposed for different types of learning problems such as supervised, self-supervised, semi-supervised, unsupervised, and reinforcement learning. Deep Learning models maintained state-of-the-art results in many fields such as computer vision, medical imaging, speech recognition, robotics, and natural language processing. The paradigm of deep learning is so expanded that, there emerged numerous models that are different than each other in many aspects. We can give examples to these models such as Artificial Neural Networks (Fully Connected Networks), Convolutional Neural Networks, Autoencoders, Recurrent Neural Networks, Spiking Neural Networks, Long Short Term Memory Networks, Gated Recurrent Units, Transformers, Variational Autoencoders, Generative Adversarial Networks, Neural Turing Machines and so on. We briefly describe Artificial Neural Networks, Convolutional Neural Networks, and Generative Adversarial Networks in this section, which are related to our work.

2.1.1 Artificial neural networks

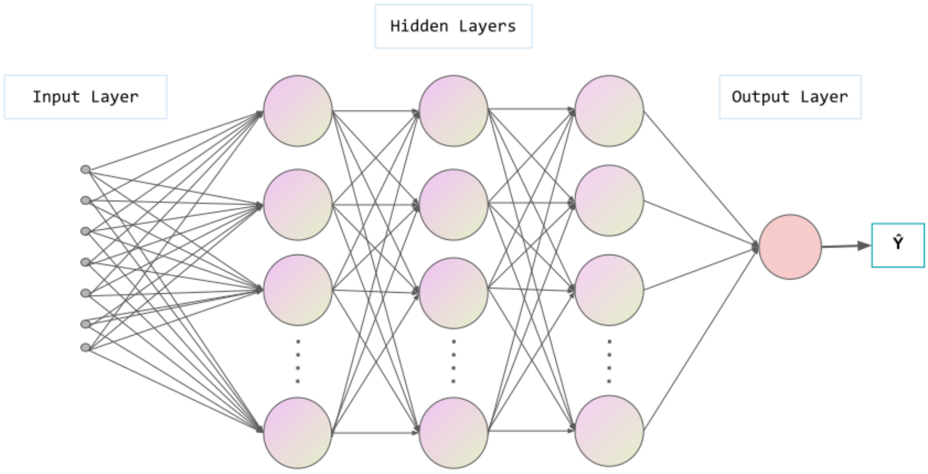

Figure 2.1 : Model of Artificial Neural Networks
One layer is constructed by following operations,
$f(x_i, W, b) = \sigma(Wx_i + b)$, W indicates weights (connections), x indicates inputs to connections, b is the bias term. σ denotes sigmoid function, where $\sigma(x) = 1/(1+e^{-x})$.

Although deep learning emerged in this decade, artificial neural networks are first designed in the 20th century. Artificial neural networks were known as multilayer perceptrons back then. They are trained with the backpropagation method. They did not become popular until recent advances in computation, developments in neural network architectures, and the existence of big data.

Artificial Neural Networks(ANN) are inspired by brain neurons, they are a rough model of the neural system of the human brain which contains neuron cells and synaptic connections between cells. As seen in Figure 2.1, circles represent neurons, and arrows represent connections between neurons. Information from neurons are accumulated through feed-forward layers using multiplication and summation operations. A nonlinearity is added after those operations in order to increase the model's representational strength. After forward-propagation, an error is calculated between the output and the ground truth, and an error signal is provided to previous layers using gradients, which is called backpropagation. In the training phase, these forward and backward operations applied on a dataset for many iterations. With enough data and a decent model, the neural network adapts itself for the successful learning of the task.

2.1.2 Convolutional neural networks

Convolutional Neural Networks (CNN) are advanced versions of ANNs. In ANNs, the existence of full connections between every consecutive layer causes a computational burden. In order to decrease the number of parameters, CNNs use shared parameters that utilize translational invariance and local neighborhoods. There are different types of layers in CNNs. The most essential layer is the convolution layer. In a convolution layer, a filter slides through an input image and it generates an output image by multiplying input pixel intensities with its weights and sums them. Another essential layer of CNNs is the activation layer, which a creates non-linearity in the network. Without activation layers, no matter how many convolution layers we add, we obtain a linear overall function. The non-linearity is essential for neural networks because it increases the representation capacity. We can give examples to various activation layers such as Sigmoid, Tanh, Rectified Linear Unit (ReLU), Leaky ReLU, and Exponential Linear Unit. Another layer that is mostly used in classification models is the pooling layer. It makes the model more robust to translational invariance. It also reduces the size of the output which causes the model to have fewer parameters in total. Pooling layers are not learnable, they do strict statistical operations such as taking the maximum value in Max Pooling and taking the average value in Average Pooling. The normalization layer is another widely used layer in CNNs, it aids the optimization process against an "internal covariate shift" during the training phase. The most used normalization method is the batch normalization, however, alternative techniques such as instance normalization, group normalization, and spectral normalization are also proposed. The training process of CNNs is similar to that of ANNs: it consists of the forward-propagation, the backpropagation, and the optimization step.

The VGG16 model is designed by Oxford Visual Geometry Group in 2014 [18]. It was the second-best model in ILSVRC 2014 competition (after GoogLeNet). As seen in Figure 2.2, it has a basic architecture that made it popular although it has too many parameters. It consists of only convolution, rectified linear unit, max pooling, fully connected, and softmax layers.

Later, it was found that the going deeper in VGG model was not giving better results after a certain amount of layers. Optimization of a deep VGG model was a difficult task particularly due to the vanishing gradients problem, and the residual connections

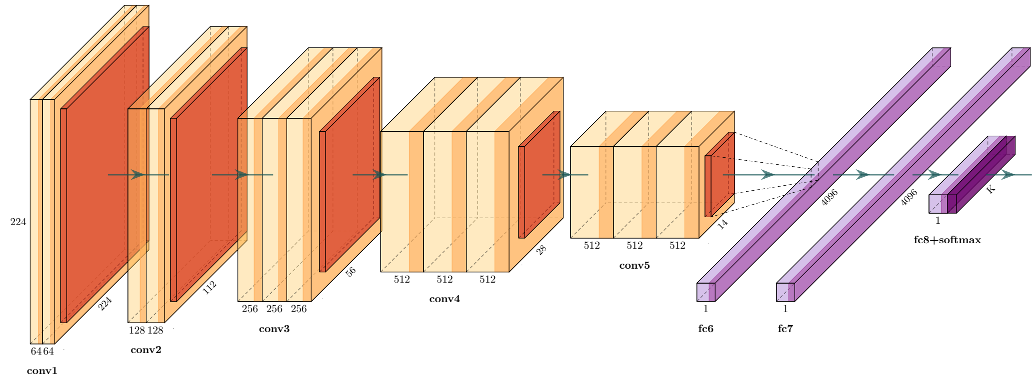

Figure 2.2 : Architecture of VGG16 model
Yellow boxes show convolution layers, red boxes show max pool layers, purple boxes show fully connected layers and dark purple box shows softmax layer.

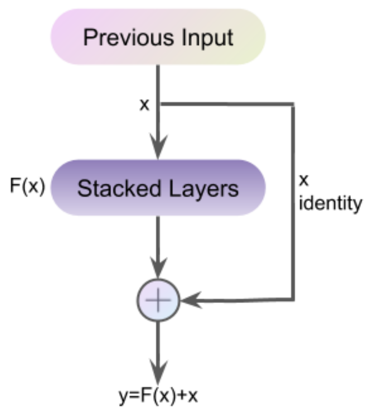

Figure 2.3 : Residual Connection.

between stacked layers fixed this issue [19]. Although they consist of a simple operation of adding a skip connection as depicted in Figure 2.3, residual layers made it possible to train CNNs with more than 100 layers. We can give examples to these networks such as ResNet101 and ResNet152 [19].

On the other hand, the UNet Architecture is designed for the semantic segmentation task in biomedical imaging [20]. Figure 2.4 presents an example for UNet style architectures. The key operation that makes UNet powerful is that it appends output of each encoder layer to the corresponding decoder layer (presented with horizontal

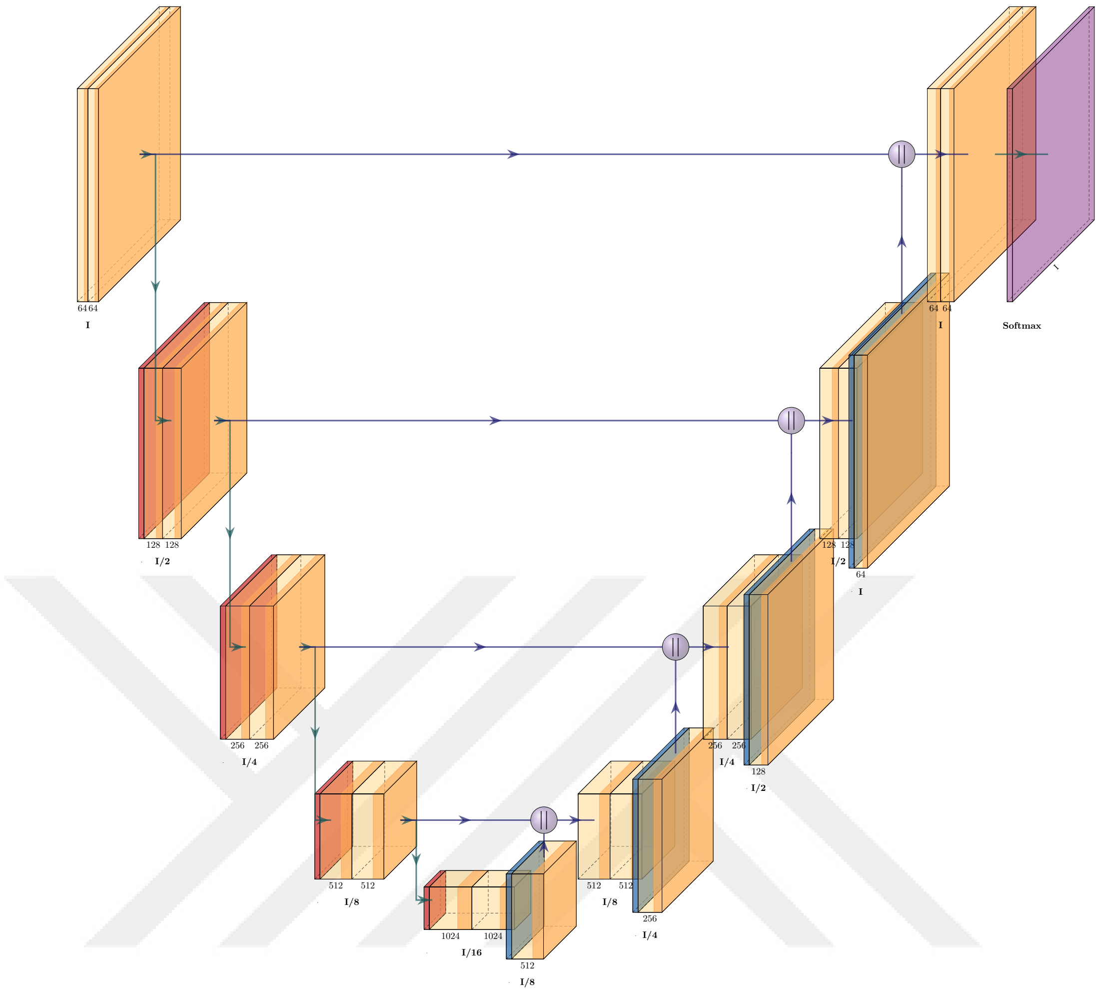

Figure 2.4 : UNet Architecture
Yellow boxes show convolution layers, red boxes show upsampling and downsampling layers, purple box shows softmax layer.

arrows). Like the residual connection, this operation also makes optimization of the training phase easier for the CNN model.

2.1.3 Generative adversarial networks

2.1.3.1 Adversarial loss (GAN - RaGAN)

Generative Adversarial Networks (GANs) belong to the class of generative networks that learn to synthesize images with a target distribution by competition of typically two networks, where one is the generator and the other one is the discriminator [2]. In vanilla GANs, as shown in Figure 2.5, the generator G learns to transform a random noise distribution to the target image distribution. Discriminator D aims to correctly classify the output of G with a "generated" label versus "real" label. Here, the "real"

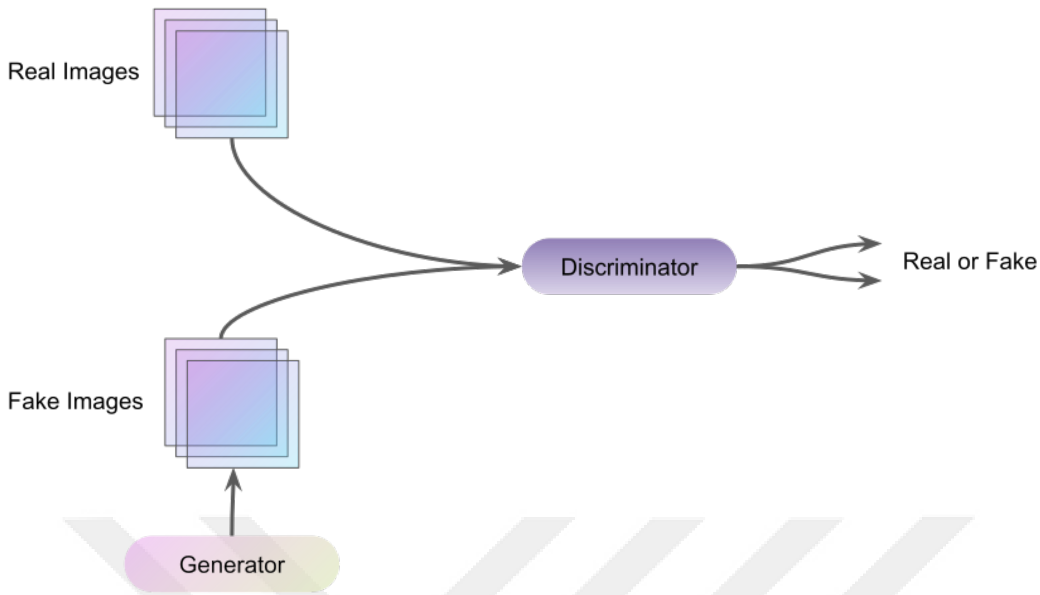

Figure 2.5 : Generative Adversarial Networks.

refers to a label of the training data. D also performs the same operation on images generated by the model. This is the basis of the adversarial loss in the vanilla GAN [2], which is used in update of both G and D:

$$L^{D}_{GAN} = -\mathbb{E}_{x_r \sim \mathbb{P}} \log(D(x_r))] - \mathbb{E}_{x_f \sim \mathbb{Q}}[\log(1 - D(x_f))] \quad (2.1)$$

$$L^{G}_{GAN} = -\mathbb{E}_{x_f \sim \mathbb{Q}}[\log(D(x_f))]. \quad (2.2)$$

Here, $D(x) = \sigma(C(x))$, where $C(x)$ refers to the final output of the discriminator network after which the activation function σ is applied. x_r refers to real data samples obtained from the dataset and x_f refers to data which is generated with generator G. A more recent GAN framework, Relativistic Average GAN (RaGAN) [21], utilizes the following losses instead:

$$L_{RaGAN} = -\mathbb{E}_{x_1}[\log(\overline{D}(x_1, x_2))] - \mathbb{E}_{x_2}[\log(1 - \overline{D}(x_2, x_1))] \quad (2.3)$$

$$L^{D}_{RaGAN} = L_{RaGAN}(x_f, x_r), \; L^{G}_{RaGAN} = L_{RaGAN}(x_r, x_f) \quad (2.4)$$

where $\overline{D}(x_1, x_2) \triangleq \sigma(C(x_1) - \mathbb{E}[C(x_2)])$. While a discriminator in vanilla GAN predicts how realistic an image is, relativistic discriminator evaluates the realness of real and fake images relatively. As it has been shown that using RaGAN loss provides sharper details while having more stable training, we also incorporate RaGAN loss in our model.

2.1.3.2 GAN models

GANs are first created with fully connected layers in Goodfellow et al's work [2]. Radford et al used only convolutional layers in their GAN model and called it Deep Convolutional GAN (DCGAN) [22]. They used convolution layers with stride 2 instead of using max-pooling layers. They added batch normalization layers in both the generator and the discriminator and used leaky ReLU in the discriminator. DCGAN has an unconditional training setting like the first GAN, which means they generate images from randomly sampled latent variables.

The Conditional GAN model is created by Mirza and Osindero in 2014 [23]. In unconditional GANs, there is no control mechanism on the generation process, only variables we manage are latent variables that do not have meaning until they are mapped to the output with the learning procedure. In the conditional GAN setting, the class label is given with a latent variable in order to control the class of the generated image, which also aids in fixing the mode collapse problem [23].

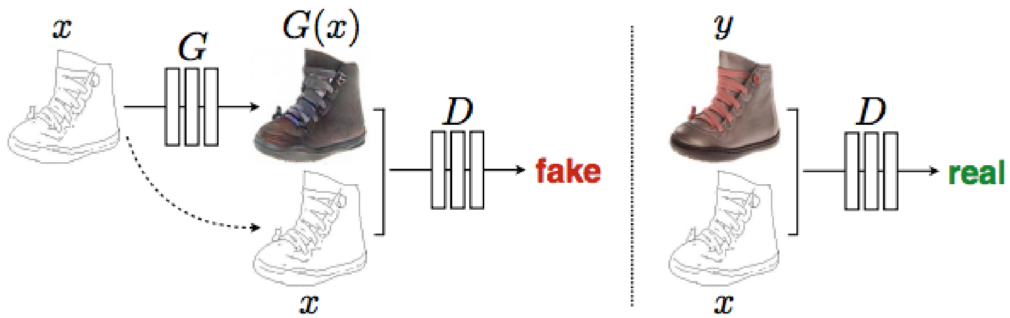

Figure 2.6 : Pix2Pix Model [1].

Image-to-Image Translation networks are special types of conditional GANs. Pix2Pix model is the first and the most popular model known in these types of networks [1]. As shown in Figure 2.6, instead of obtaining class labels as conditions, the model takes an input image as a condition and generates an image according to that input image. Discriminator gets an input image of the generator and generated image concatenated together as a fake batch. Real batch also consists of an input image and ground truth image together. Image-to-Image Translation networks are widely used in translation tasks such as translation of grayscale image to colorized image, low-resolution image to high-resolution image, aerial image to map, day to night, edges to photo, semantic labels to scene image as seen in Figure 2.7. Pix2Pix model also provides a PatchGAN architecture for its discriminator, which means it creates one realness value for each receptive field instead of one value for the whole image.

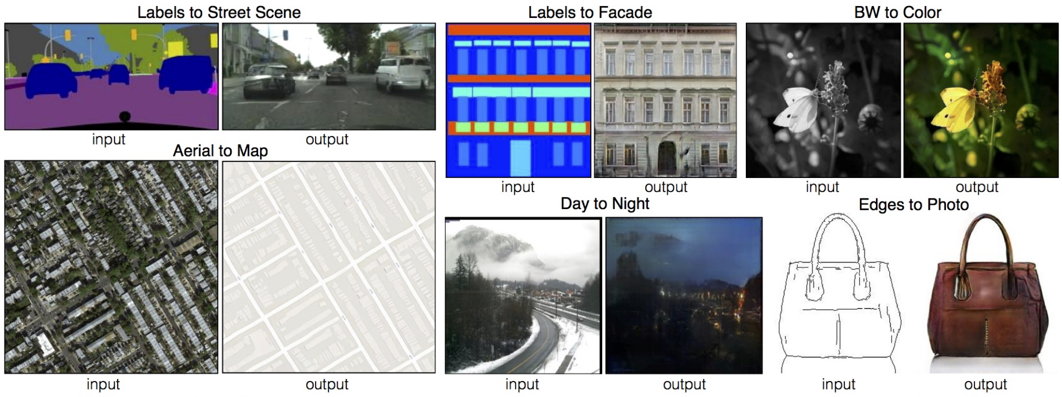

Figure 2.7 : Example of Image-to-Image Translation tasks [1].

2.2 Related Self-Supervised Learning Tasks on Images

As we mentioned in Chapter 1, CNN-based pansharpening methods that are maintained before this work, utilized super-resolution task for the training framework and we have proposed a framework inspired by the colorization task. We further define these self-supervised tasks in this section.

2.2.1 Image super-resolution

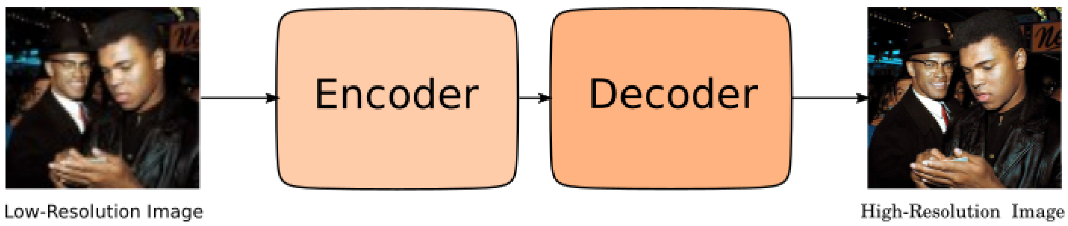

Figure 2.8 : Super-resolution network scheme.

Image super-resolution is about enhancing the resolution and the perceptual quality of the image [3]. In the training phase, as shown in Figure 2.8, original images are downsampled with bilinear or bicubic interpolation and upsampled back again in order to obtain low-resolution images although numerically they have the same size as the ground truth images. These low-resolution images are given to the convolutional neural networks as model inputs. The outputs of the model are the predicted high-resolution images that have the same resolution as the ground truth images in the training phase.

The loss is calculated by the mean squared or absolute error between the generated and the ground truth images pixel-wise. In real-life testing, we can directly upsample images (without downsampling first) then provide them to the CNN in order to get high-resolution images.

2.2.2 Image colorization

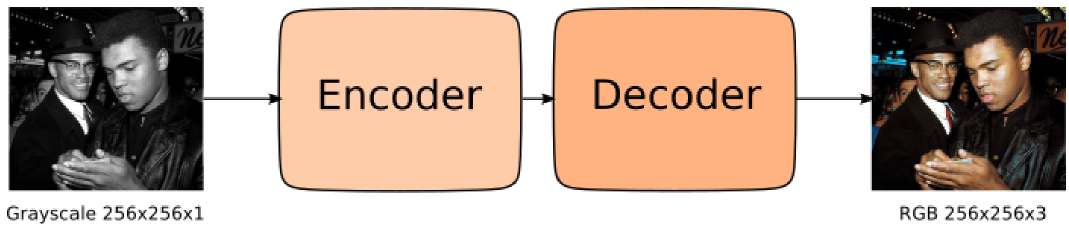

Figure 2.9 : Colorization network scheme.

Image colorization is about colorizing the grayscale images [24]. In the training phase, as shown in Figure 2.9, images with color are transformed to grayscale using a weighted sum of RGB channels and they are given as input to the CNN model. Outputs of the model are images with 3 channels which have the same size as ground truth images. The more conventional way of doing colorization is that transforming RGB to LAB color space first and providing L, the lightness channel as an input, and predicting AB, the color channels [24]. The loss is calculated by the mean squared or absolute error between the generated AB channels and the AB channels of ground truth images. In real-life testing, we can provide images which originally do not have a color like photos from before 70's.

2.3 Pansharpening

2.3.1 Traditional pansharpening methods

Traditional methods of pansharpening algorithms can be separated mainly into two categories: component substitution based methods and multiresolution analysis methods [25]. Component Substitution (CS) methods transform and split multispectral images into spatial and spectral components, then try to replace the spatial component with a component obtained from panchromatic images. Many variants of CS methods such as PCA, IHS, GS, Brovey Transform, BDSD, and PRACS appeared in the literature [26]. Multiresolution Analysis (MRA) methods mainly obtain spatial information by first applying a filter to panchromatic images, followed by an injection of the obtained information to multispectral images [27]. There are many examples of MRA methods such as the high-pass filtering (HPF), MTF based methods like Generalized Laplacian pyramids with modulation transfer function (MTF-GLP), MTF-GLP with high pass modulation (MTF-GLP-HPM), MTF-based algorithms with spatial principal component analysis (SPCA) and wavelet-based methods like ATWT, UDWT, and AWLP [28–33].

2.3.2 CNN-based pansharpening methods

Similar to many image processing tasks, CNNs are also used for pansharpening after their success in image synthesis. For the training process of CNNs for the pansharpening task, Wald's protocol is utilized [15]. Since no reference images exist in full-resolution pansharpening, a reduced-resolution setting is used for the training. This is similar to what is done in the super-resolution task. Masi et al suggested a three-layer CNN model that is inspired from super-resolution using deep convolutional neural networks [6]. Yang et al have proposed a model which uses knowledge specific to the domain to enhance structural and spectral properties, while employing high-pass filtering instead of using directly the image [7]. Huang et al used a stacked modified sparse denoising autoencoder for pretraining a deep neural network model effectively [8]. Liu et al established a model that fuses information gathered from panchromatic and multispectral images at a feature level after several convolution operations [10]. Later, they enhanced the model via a generative adversarial framework by adding a discriminator network [11]. Scarpa et al utilized a pretrained model that does a fine-tuning on the target image before the inference stage [12]. Wei et al designed a convolutional neural network that uses deep residual learning [14]. In a recent study, Vitale et al devised a cross-scale learning model where it combines losses from both reduced resolution and full resolution comparisons [34]. Although there are many variants of CNN-based pansharpening models, they follow a common framework that has major limitations, as we discuss in the next chapter.

3. MOTIVATION

In this chapter, we elucidate issues with the super-resolution based pansharpening approach. First, we describe the standard CNN-based pansharpening framework that is inspired by the super-resolution task in Section 3.1. In Section 3.2, we present the spatial detail differences across reduced resolution panchromatic images and full resolution multispectral images. We also demonstrate why current pansharpening with deep learning approaches are not efficiently handling this problem in the same section. In Section 3.3, we discuss the blurring problem that is caused by an inherent uncertainty in the ratio between full resolution and reduced resolution images.

3.1 Standard CNN-based Pansharpening Framework

As stated in Chapter 2, several pansharpening models were built on CNNs or GANs in the recent literature. Although they offer various architectures, their underlying learning procedures are similar. The standard procedure in CNN-based pansharpening methods that utilize the Wald's protocol, which is designed to overcome the reference problem in quantitative analysis of pansharpening. In Wald's protocol, the algorithm gets the reduced resolution panchromatic image and the reduced resolution multispectral image as input, and tries to produce an image similar to the original multispectral image as its output through various image processing operations. Deep learning-based models, on the other hand, involve extensive training processes that are designed while adopting Wald's protocol.

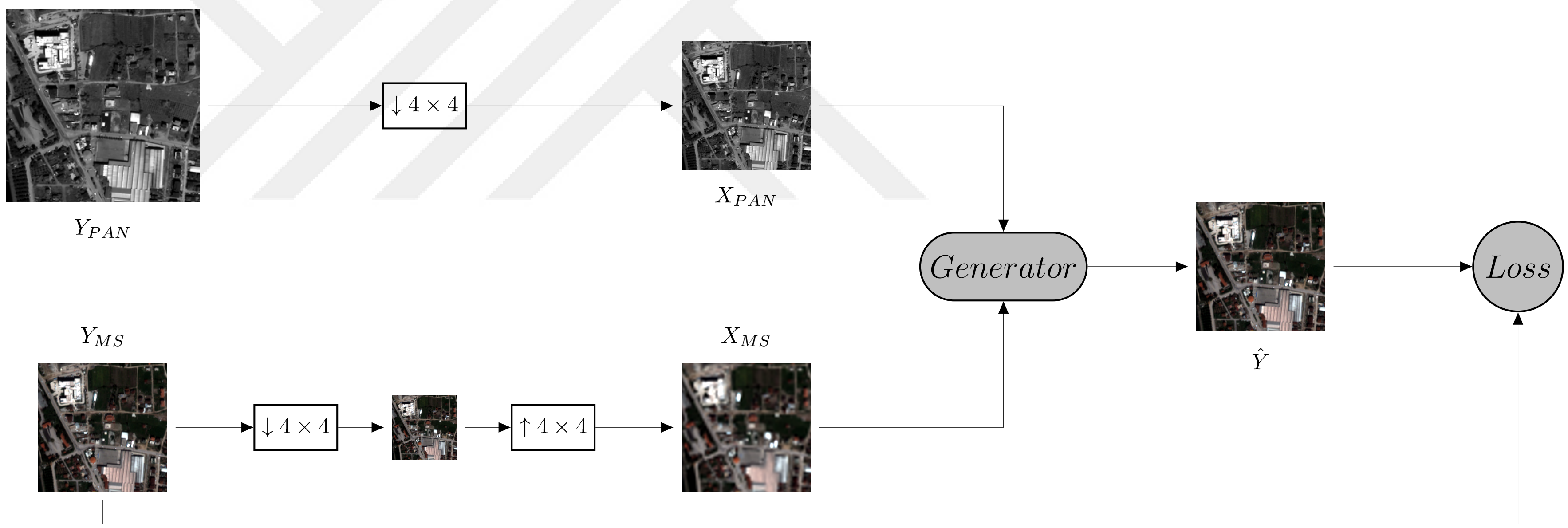

Figure 3.1 : Standard framework for CNN-based pansharpening. Generator refers to a CNN model. A loss function is calculated between $\hat{Y}$ and Y_{MS} in order to train the CNN model. Generator learns to produce the pansharpened image $\hat{Y}$ using loss signals.

Table 3.1 : Quantitative analysis of spatial quality incompatibility between reduced panchromatic images and multispectral images.

	PSNR	sCC	SSIM
(worst-best)	(0-inf)	(0-1)	(0-1)
Reduced PAN - Grayscale MS	24.704	.088	.586
Reduced PAN(Blurred) - Grayscale MS	30.751	.424	.848

We illustrate the standard CNN-based pansharpening framework in Figure 3.1. Suppose that we have Y_{PAN} and Y_{MS}, which are corresponding panchromatic (PAN) and multispectral (MS) images that we want to fuse through pansharpening. First, Y_{PAN} is reduced by 4$\times$ to the size of the Y_{MS} to obtain the X_{PAN} image. Y_{MS} is reduced by 4$\times$, then upsampled by 4$\times$ to obtain the X_{MS}. X_{PAN} and X_{MS} are provided to a generator network G, hence $\hat{Y} = G(X_{PAN}, X_{MS})$ is obtained at the output as the generated or pansharpened image. A reconstruction loss function, either with an L_2 or L_1 norm is measured between the multispectral image and output.

The procedure with standard CNN-based models with or without an adversarial loss then is executed through an optimization of the overall loss function (see Section 2.1.3). Next, we explain the disagreement in spatial details after training such a model.

3.2 Problems in Reduced Resolution Pansharpening

When one trains a model with the standard CNN-based pansharpening framework, although quantitative results between original multispectral and generated pansharpened images are typically highly favorable, a closer inspection of the inputs and outputs shows that pansharpened images that are obtained from the model do not preserve the desired spatial information that presents sharp details in the reduced panchromatic image inputs. We notice that the problem lies within the crucial assumption that the reduced panchromatic images and original multispectral images should have similar spatial quality as they bear the same spatial resolution level. On the contrary, it can be both qualitatively and quantitatively argued that the reduced panchromatic images exhibit better spatial quality than original multispectral images.

Figure 3.2 qualitatively demonstrates this problem, where spatial detail disagreement in terms of lack of sharpness in detail, blurriness, reduced contrast differences, and less continuity in lines in the images can be clearly seen by visual inspection (compare

(a)

(b)

(c)

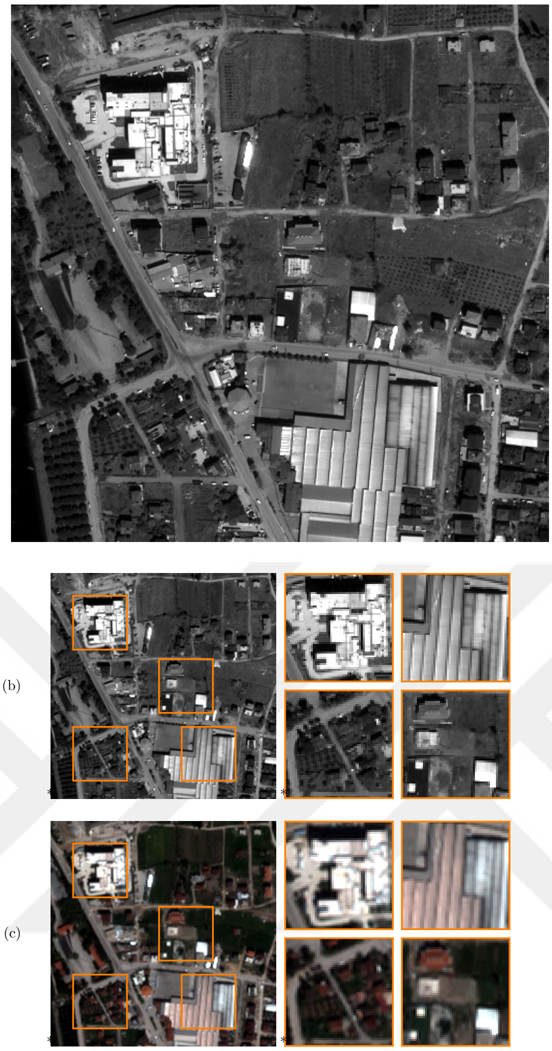

Figure 3.2 : Spatial-level-of detail comparison between reduced panchromatic and multispectral images demonstrated on Pleiades dataset. (a) Original panchromatic image. (b) Reduced panchromatic image. (c) Multispectral image. Orange boxes on the left are zoomed into for display on the right.

zoomed image patches in (b) and (c)). In order to quantitatively test our conjecture, we calculate three measures, which are PSNR, sCC, and SSIM on a set of reduced panchromatic images given the corresponding gray-transformed multispectral images as a reference. Next, we apply a blurring Gaussian filter with 5×5 kernel ($\sigma = 2$) to obtain the blurred reduced panchromatic images. We calculate the three measures using this time the blurred panchromatic image rather than the original panchromatic image (Table 3.1). Per our hypothesis, quantitative measures should improve with blurred versions of the reduced panchromatic images, since we claim that original multispectral images are blurrier than reduced panchromatic images. It can be observed in Table 3.1 that all three measures change in an expected direction, hence the blurred versions of the reduced panchromatic images show increasingly similar characteristics to multispectral images.

Current deep learning methods used in pansharpening, which are inspired mainly from super-resolution, inherently incorporate the abovementioned spatial detail disagreement issue into their procedures as they involve mapping a function from a pair of reduced resolution panchromatic image and reduced resolution multispectral image to the original multispectral image. Our analysis above shows that reduced resolution panchromatic images contain more spatial details than the original multispectral image, which are lost during the prescribed procedure. This is the main reason behind obtaining decent quantitative results, whereas pansharpened images exhibit reduced spatial details compared to original panchromatic images.

3.3 Problems in Full Resolution Pansharpening

Similar to the reduced resolution procedures, the full-resolution pansharpening procedure is also prone to a specific blurring problem due to the strong assumption of learning a “fixed upsample scale” (e.g. say a typical ratio of $4\times$) in the training phase of standard CNN-based approaches. As the level of detail of the $4\times$ reduced resolution panchromatic image can not match to that of the corresponding original multispectral image, training the CNN-based learning model according to the Wald’s Protocol naturally cannot match the desired upsampling ratio exactly, and leads to blurry results for the full resolution case. We present a remedy to that problem, by introducing random downsampling ratios, rather than a fixed (e.g. $4\times$) reduced scale during training, as the latter does not generalize well to full resolution pansharpening, as is demonstrated in Chapter 5.

4. PANSHARPENING WITH GUIDED COLORIZATION USING GANS

To address the shortcomings of the standard CNN-based approaches, we present a new pansharpening method that faithfully preserves spatial details given by the input panchromatic image in the inference stage. This is achieved by designing a self-supervised learning procedure based on the colorization task rather than super-resolution task. This new task that is cast upon the network model requires that during the training phase, we provide the grayscale multispectral image, whose spatial details perfectly agree with those of the original multispectral image. This is not the case for the reduced panchromatic image due to spatial detail disagreement problem that we discussed in Section 3.2.

To further expound our reasoning on colorization based pansharpening versus super-resolution based pansharpening, an analogy of comparison between traditional CS and MRA methods can be made. Existing super-resolution based pansharpening methods can be considered more similar to MRA methods than CS methods because, in the training phase, the model tries to increase spatial details of reduced resolution multispectral image with spatial features extracted from reduced resolution panchromatic image by comparing it to the original multispectral image. On the other hand, the colorization based pansharpening method we propose can be interpreted more in line with a CS approach rather than an MRA approach. As we will see more details in the following parts, our model learns to generate an original multispectral image by taking its reduced resolution multispectral image and the corresponding grayscale multispectral image as inputs, which is interpreted as colorization. We can also interpret this in a way that our model learns to separate spectral and spatial components of the multispectral image during training. Then, in the testing stage, we provide the corresponding panchromatic image instead of the grayscale multispectral image, which can be interpreted as substitution of spatial components between two images, which alludes to traditional CS approaches.

Furthermore, we improve the full-resolution pansharpening procedure by injecting noise into the assumed downsampling-upsampling ratios between the original panchromatic and multispectral images, which induces a regularization effect into our model.

The proposed PanColorGAN pansharpening learning model is illustrated in Figure 4.1. First, let us describe the original PanColorGAN with a fixed down/up-sampling ratio. Suppose that the input multispectral image Y_{MS} is first downsampled by $k = 4\times$ then upsampled by $k = 4\times$ to obtain X_{MS}. Y_{MS} is also transformed to grayscale by taking an average of channels to construct a grayscale input X_{GMS}. Later, X_{GMS} and X_{MS} are provided as input to the generator network G and $\hat{Y}_G = G(X_{GMS}, X_{MS})$ is obtained as the output. A reconstruction loss is calculated between $\hat{Y}_G$ and Y_{MS}.

In our PanColorGAN, as in traditional GANs, an additional discriminator network D is also built to provide an Adversarial Loss, which is calculated for $\hat{Y}_G$ because we would like to augment the representation capability of the generator network by providing feedback on the quality or the credibility of its generated output. Details of the model are explained next.

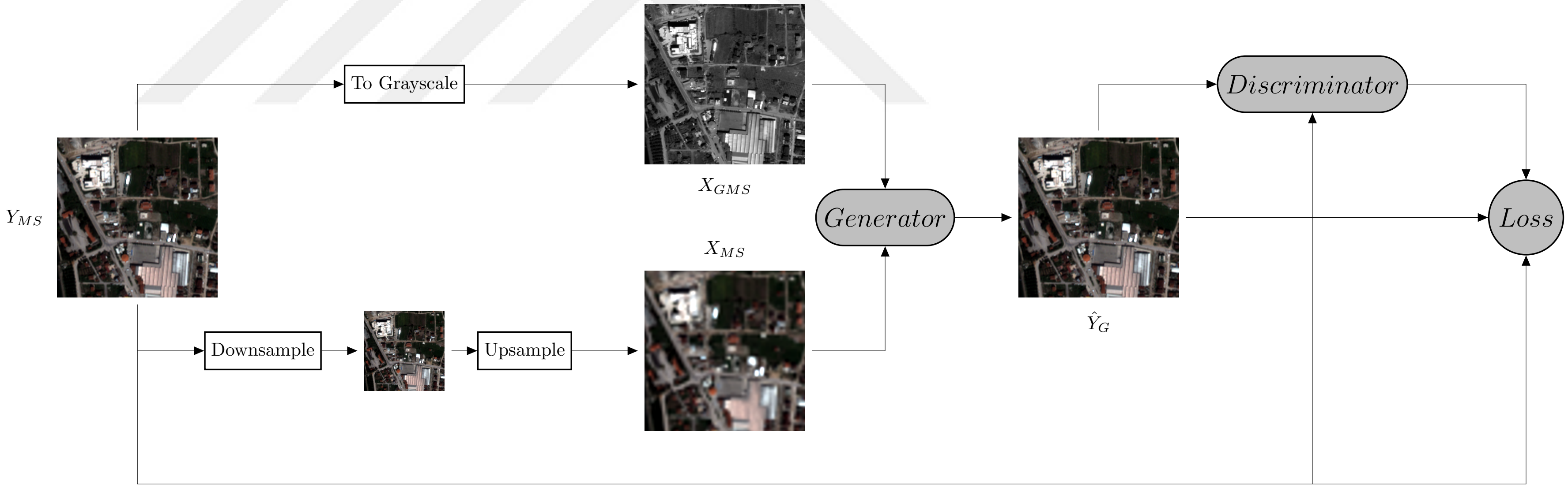

Figure 4.1 : Proposed training scheme for PanColorGAN model: A reconstruction loss $Loss(L_1)$ between the colorized output of the $\hat{Y}_G$ input and Y_{MS}, as well as an adversarial loss that evaluates the generation quality of $\hat{Y}_G$ generated from X_{GMS} and X_{MS} are utilized to train the PanColorGAN.

4.1 PanColorization GAN (PanColorGAN) Model

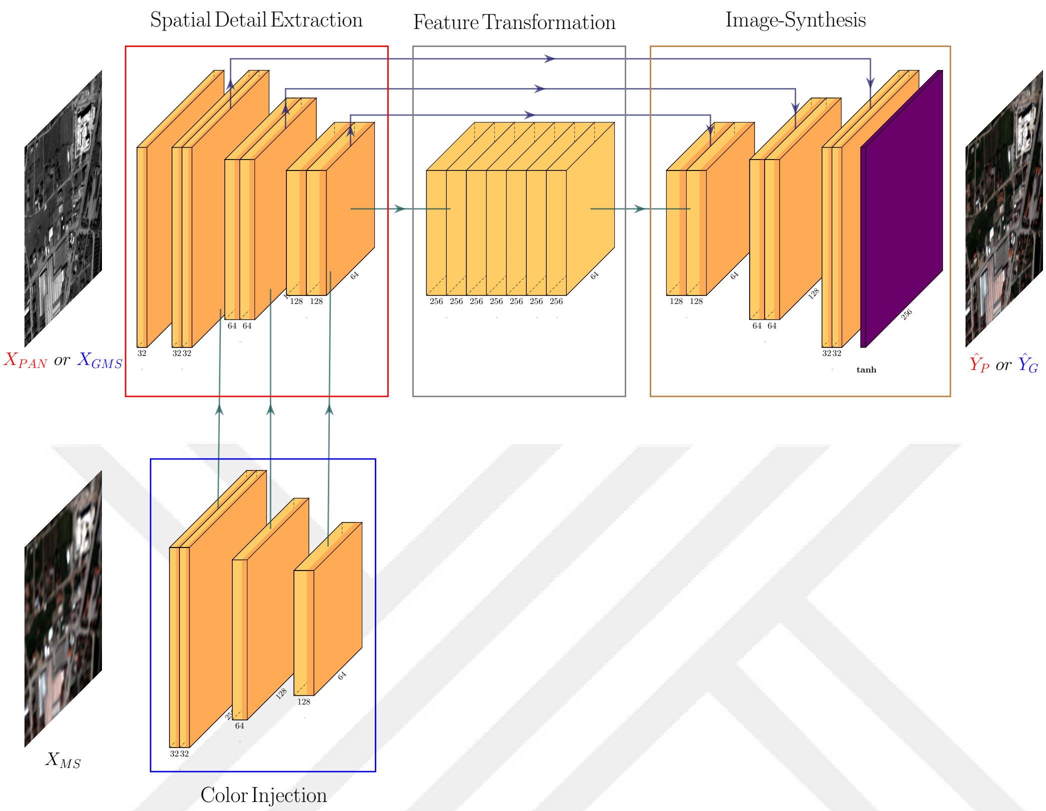

Figure 4.2 : Generator of PanColorGAN model: Architecture details for its Generator network are depicted. Two modes exist for Generator network: In the training phase, X_{GMS} is provided along with X_{MS} to generate $\hat{Y}_G$. In the testing phase, X_{PAN} is provided along with X_{MS} to generate $\hat{Y}_P$.

Figure 4.2 depicts the details of the Generator of PanColorGAN architecture. Its generator G is a modified and expanded version of the UNet [20] architecture. It has shortcuts of concatenation across layers in order to provide improved optimization in terms of reducing the vanishing gradients problem. G has four main parts that serve specific goals: (i) spatial detail extraction, (ii) color injection, (iii) feature transformation, and (iv) pansharpened image synthesis. The spatial detail extraction part takes a grayscale image (X_{GMS}) as input and applies 3×3 convolutions while obtaining color features from the color injection part. The color injection part is a fully convolutional architecture that applies 3×3 convolutions four times and injects extracted color features from the multispectral image (X_{MS}) to spatial detail extraction layers of the network after every convolution operation except the first one. There is a residual block in the middle of the network that transforms concatenated spatial

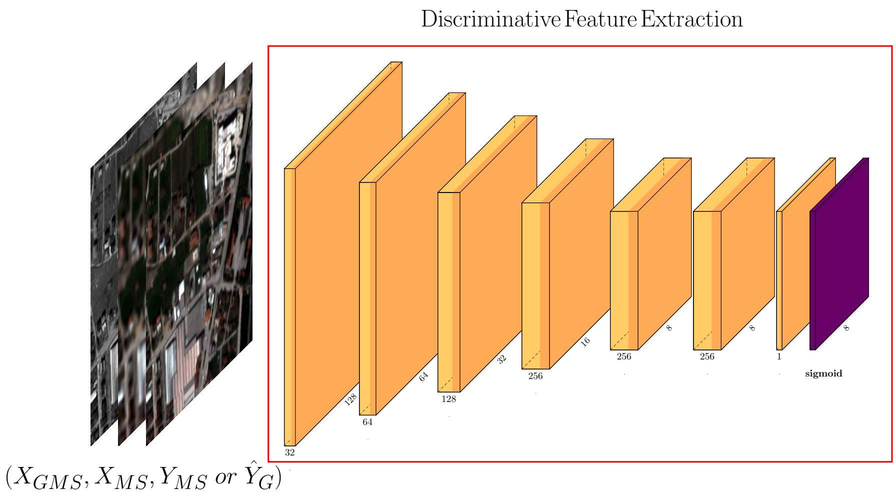

Figure 4.3 : Discriminator of PanColorGAN model: Architecture details for its Discriminator network are depicted. During the training phase, Discriminator network gets two different types of batches. A real batch consists a concatenated set of X_{GMS}, X_{MS} and Y_{MS}. A fake batch consists a concatenated set of X_{GMS}, X_{MS} and $\hat{Y}_G$, as shown on the bottom right.

and spectral features and prepares them for a synthesis of the pansharpened image. Finally, the network slowly increases height and width, and decreases the depth of features by applying upsampling and 3×3 convolutions while obtaining features from the detail extraction part, as in the standard Unet architecture. Batch normalization and LeakyReLU activation are inserted after every convolution operation. After obtaining features as the same dimension as the multispectral image, the tanh activation is applied to map the image intensities to [-1,1] interval. Using tanh provides faster and more stable training of GANs [22]. This produces the output $\hat{Y}_G$ of the generator network.

Figure 4.3 depicts the details of the Discriminator of PanColorGAN architecture. PanColorGAN discriminator D has a conditional patchGAN architecture [1], which operates on image patches, and gives an output for every receptive field it sees. Hence, the output indicates whether those receptive fields seen by D look realistic or not. Then those outputs are aggregated in a patchGAN-loss for the training of the discriminator network D. The reconstruction loss for the training of the Generator is not calculated over patches, but calculated pixel-wise for the whole image. In a conditional GAN framework, conventionally the D network takes the generated image from the generator network or ground truth image along with inputs. Pansharpening can be regarded in the framework of image-to-image translation idea, which was first presented in the study of Isola et al [1]. In image-to-image translation with conditional adversarial networks, for the discriminator network D, the inputs to the generator are taken as conditions in its decision of "real" or "fake". For that reason, during our training, fake batches consist of grayscale images X_{GMS}, reduced multispectral images X_{MS}, and outputs of G network $\hat{Y}_G$. Real batches consist of X_{GMS}, X_{MS} and original multispectral images Y_{MS}. This procedure differs from that of the unconditional generative adversarial networks where the generator network synthesizes images from randomly sampled latent variables and the discriminator receives only the generated images and real images at its input. Providing all related inputs with generated and real images ensure that the discriminator network understands visual relations between input and output images. D applies 4×4 convolutions with 2-strides 5 times and reduces height and width while increasing depth. Then a final convolution reduces the depth to 1. Batch normalization and LeakyReLU activation are executed after every convolution layer. Sigmoid operation is applied in order to shrink the interval to [0,1]. Hence, at the output of D, indicators of the realness of receptive fields in the given image are obtained.

PanColorGAN model utilizes the following losses for learning the weights of the G and D networks:

$$L_D = L_{RaGAN}(Y_{MS}, \hat{Y}_G) \tag{4.1}$$

$$L_G = L_{Rec} + \alpha L_{RaGAN}(\hat{Y}_G, Y_{MS}) \tag{4.2}$$

$$L_{Rec} = \| Y_{MS} - \hat{Y}_G \|_1 \tag{4.3}$$

The loss of reconstruction is described as the mean absolute error (L1 loss), whereas the adversarial loss is designed as the relativistic average GAN loss. While the reconstruction loss increases pixelwise similarity between generated images and corresponding multispectral images, the adversarial loss brings closer the distribution of generated images to multispectral images and provides sharpness in detail. In PanColorGAN, L_{Rec} measures the distance between Y_{MS} and $\hat{Y_G}$ rather than $\hat{Y_P}$, because the latter would lead the training network to bias the spatial distribution of the pansharpened image towards the multispectral image domain, which is not ideal, as argued before in Chapter 3.

4.2 Random Downsampling of Multispectral Images

As we discussed in Section 3.3, training the pansharpening network with $4\times$ downsampling scale reduces the representation capacity of the model, particularly for the full resolution pansharpening scenario. Hence, we substitute $4\times$ downsampling operation with a random downsampling operation in an enhanced model, which we call PanColorGAN+RD (Random Downsampling). As we want the model to learn the colorization of grayscale transformed multispectral images and panchromatic images, the model should be robust to variations in the spatial resolutions of the reduced multispectral images, which are used for their spectral information. When random downsampling procedure is used for an image, say with height and width sizes of 256, instead of downsampling the image to a fixed size of 64×64, we sample an integer, say s, from a uniform random distribution between (a, b), where a and b are two predefined numbers (See Section 5.1). We downsample the image to the selected size $s \times s$, and then immediately upsample it back to 256×256. We emphasize here that this random downsampling process is applied only during the training phase of the network. In the testing phase, random downsampling is not utilized. This modification provides a way to PanColorGAN to improve its learning as follows: when only $4\times$ downsampling is used in the training stage, the network learns to interpolate the reduced panchromatic image and the reduced multispectral images with the given scale and does not learn the colorization task properly. As the actual spatial resolution scale difference between the former two is not known exactly, the learned result provides neither the desired nor the sufficient super-resolution level when the model is applied on full resolution. This effect is demonstrated in Section 5.4.

4.3 Inference through proposed PanColorGAN models

After the training phase is completed, during the reduced resolution testing phase, the original Y_{PAN} image is reduced to the same size as the multispectral image to obtain X_{PAN}. The X_{PAN} and X_{MS} images are provided to the trained PanColorGAN generator network G and $\hat{Y} = G(X_{PAN}, X_{MS})$ is obtained as the output, for the reduced resolution inference.

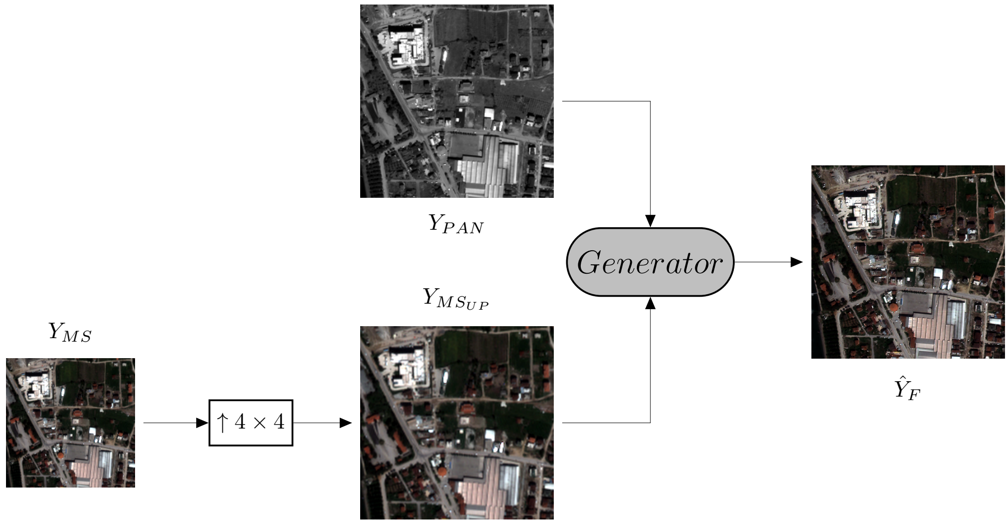

Figure 4.4 : Full resolution inference (testing) scheme: Y_{MS} is upsampled by $4\times$ to obtain $Y_{MS_{UP}}$. $Y_{MS_{UP}}$ and Y_{PAN} are fed to the trained PanColorGAN generator in order to get the full resolution pansharpened image $\hat{Y}_F$ at the output.

Figure 4.4 illustrates how to execute the full resolution, i.e. the real life scenario in pansharpening. The original Y_{PAN} and $4\times$ upsampled version of Y_{MS} are provided to the trained PanColorGAN generator network G, and $\hat{Y}_F = G(Y_{PAN}, Y_{MS_{UP}})$ is obtained as the full-resolution pansharpened image output.

5. EXPERIMENTS AND RESULTS

In this chapter, we provide implementation details of experiments, utilized datasets and evaluation indexes, quantitative and qualitative evaluation of reduced resolution and full resolution results. Furthermore, we present transferability properties of our model as well as discussions of the results.

5.1 Implementation Details

We implemented PanColorGAN in Pytorch 1.0 and trained it on one Titan RTX GPU. An iteration in the training phase takes approximately 2 seconds, which makes an epoch approximately 1 hour for our training set. We trained our models for 100 epochs and selected the best checkpoint in the latest epochs in terms of performance, which took a model 4 days to train. As a baseline GAN-based pansharpening method, we build a pansharpening model inspired by the super-resolution task, which is similar to other standard CNN-based methods. We name it as PanSRGAN, which is trained with X_{PAN} input instead of X_{GMS}, following the same procedure in standard CNN-based pansharpening framework. We compare it with our PanColorGAN model in order to perform an ablation study to assess the provided improvements.

Disabling the adversarial loss and using only the reconstruction loss leads to blurrier image generation. This blurriness property occurs due to characteristics of reconstruction loss, for instance as in pixel-wise minimum squared error loss that tends to average details of local neighborhoods. Adversarial loss provides a perceptual similarity metric to training which leads to sharper results in contrast to reconstruction loss [35]. The advantages of using generative adversarial networks instead of only generators with reconstruction loss were reported in the study of Liu et al [11] for the pansharpening case as well. Considering the beneficial effects of adversarial loss in image generation, we also adapt the generative adversarial network framework to all pansharpening models proposed in this work.

Table 5.1 : Information of satellite images in datasets.

Region	Satellite	MS/PAN m	Across Track	Along Track	Train/Test
Aydin	Pleiades 1A	2 / 0.5	-6.91	18.12	Train
Istanbul	Pleiades 1A	2 / 0.5	-22.92	-11.15	Train
Istanbul	Pleiades 1A	2 / 0.5	4	-18.77	Train
Bursa	Pleiades 1A	2 / 0.5	4.32	-14.60	Train
Bilecik	Pleiades 1A	2 / 0.5	3.08	-13.89	Train
Mugla	Pleiades 1A	2 / 0.5	-9.08	15.73	Test
Stockholm	Worldview 2	1.6 / 0.4	6.20	-7.10	Test
Rio	Worldview 3	1.2 / 0.3	23.90	-2.50	Test
Tripoli	Worldview 3	1.2 / 0.3	-3.70	5.00	Test
Washington	Worldview 2	1.6 / 0.4	10.10	-7.70	Test

In our experiments, the mini-batch size was set to 16. We used Adam optimizer with an initial learning rate 0.0002, β_1 as 0.5 and β_2 as 0.999. We did not use weight decay because it decreased the performance of image synthesis. Adversarial loss weight α was set to 0.005 in Eq. 4.2. A leakyReLU activation with 0.2 slope is used in all activation layers. During the training of the PanColorGAN+RD model, for each image in a given batch, a random downsampling size is sampled uniformly as an integer from the $[20, 80]$ interval. The upsampling scale is then automatically set to upscale the downsampled image back to 256. Both upsampling and downsampling are carried out with a bicubic interpolation scheme.

5.2 Dataset and Evaluation Indexes

The first dataset consists of 6 full-sized image scenes from Pleiades 1A&1B twin satellites owned by AIRBUS. Five of them are used for training and one of them is used for testing. Frames are divided into patches of 1024×1024 for panchromatic images, 256×256 for multispectral images. Thus, 30000 training samples and 5700 test samples are gathered for the Pleiades dataset. Pleiades image data includes RGB channels together with near-infrared with $2m$ spatial resolution for multispectral images. Its single-banded panchromatic image has $0.5m$ resolution. The dataset consists of images from both rural and urban areas in Turkey. In addition, image acquisition angles and seasons are in a wide range, which helps to train the model with a dataset that reflects various illumination and geometric conditions. The second dataset we utilize in our testing experiments consists of four image scenes from Worldview 2 and Worldview 3 satellites owned by Digital Globe (Maxar Technologies), which is published as open source [36]. We extract 350 patches (256×256 MS, 1024×1024 PAN) from 4 cities, which are Stockholm, Washington, Tripoli and Rio. Similar to the Pleiades dataset, Digital Globe data has 4 channels for multispectral images which are RGB and near-infrared. The spatial resolution of 4-band multispectral data is 1.6m and single panchromatic data is 0.4m for Worldview 2 images, while the resolution of 4-band multispectral data is 1.2m and single panchromatic data is 0.3m for Worldview 3 images. Both Pleiades and Worldview images were obtained in UTM projection system with appropriate zones. Detailed information about the image dataset is provided in Table 5.1. We trained the following models: (i) the proposed PanColorGAN; (ii) PanColorGAN+RD: PanColorGAN with Random Downsampling; (iii) PanSRGAN: the baseline GAN-based pansharpening model; (iv) TA-CNN: Target-Adaptive CNN-based pansharpening [12]; (v) PanNet: Deep Network for Pansharpening [7]. For comparison, we also utilize traditional pansharpening algorithms that are available in the Open Remote Sensing repository [37] which are BDSD [38], ATWT [28], GSA [39], GLP-REG-FS [40], NIHS [41], and Semiblind Deconv [42]. For training TA-CNN and PanNet models, we used the codes supplied by the authors [43, 44].

For the quantitative analysis, across all algorithms including the baselines, QAVE [45], SAM [46], ERGAS [47], sCC [48], and Q [49] are used as performance measures that include references in their calculations. We also analyze all algorithms in full resolution with no-reference metrics. Non-reference performance measures we utilize are D_s, D_λ, and QNR [50]. For calculation of all metrics, again we use the Matlab™code in Open Remote Sensing repository [37].

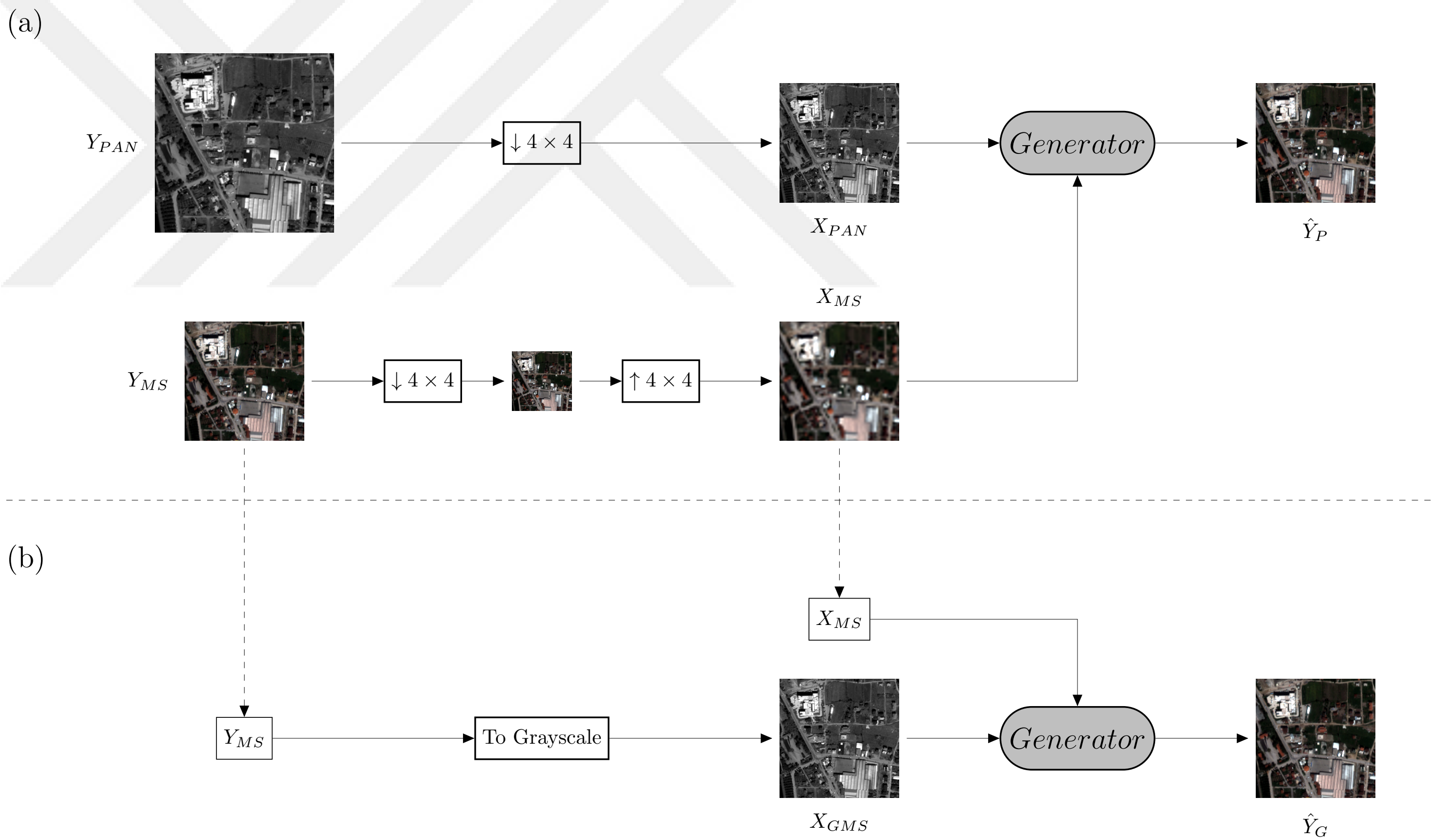

Figure 5.1 : Reduced resolution testing scheme. (a) Y_{PAN} is downsampled by 4× to obtain X_{PAN}. Y_{MS} is downsampled and then upsampled by 4× to obtain the X_{MS} image. X_{MS} and X_{PAN} are given to the generator G in order to get pansharpened image $\hat{Y}_P$ in the natural operation mode of the PanColorGAN, PanSRGAN, and other CNN-based pansharpening models. (b) This mode is shown only for evaluation of the training procedure of PanColorGAN-GMS and PanColorGAN+RD-GMS models: Y_{MS} is converted to grayscale to obtain X_{GMS}. X_{MS} and X_{GMS} are fed to the generator G to obtain the colorized image $\hat{Y}_G$.

5.3 Evaluation of Reduced Resolution Results

Figure 5.1 depicts reduced resolution testing scheme. We construct two versions of the method during inference, where we provide: (1) grayscale multispectral image alongside reduced multispectral images to obtain PanColorGAN-GMS; (2) reduced panchromatic image alongside reduced multispectral images to obtain PanColorGAN-PAN model. Similarly, two versions PanColorGAN+RD-GMS and PanColorGAN+RD-PAN models are constructed for the random-downsample version of our method. Reduced panchromatic images and reduced multispectral images are utilized for traditional pansharpening algorithms, CNN-based methods, and the PanSRGAN model.

5.3.1 Quantitative Analysis of Reduced Resolution Results

For all the with-reference measures in Table 5.2, PanColorGAN-GMS outperformed all other techniques, both CNN-learning based, and previous traditional approaches. PanColorGAN-GMS surpasses PanColorGAN-PAN extension models, where for the latter, the reduced panchromatic image is used as the input during inference. This is expected because the training procedure is set up to force the model to learn to colorize the gray-transformed multispectral image, hence the loss functions make use of the grayscaled multispectral images, not the reduced panchromatic images. Also, although standard CNN-based models such as PanNet, TA-CNN, and PanSRGAN perform clearly worse in visual quality (demonstrated later), they obtain second-tier yet close performances to other PanColorGANs, still staying behind PanColorGAN-GMS.

5.3.2 Reduced Resolution Scenario Visual Results

Figure 5.2 shows results from all algorithms on Pleiades test dataset. The corresponding full-resolution panchromatic image was given in Figure 3.2 on the left. Images in (c)-(h) belong to the results of traditional approaches, and (i)-(o) depict results of the CNN-based methods. Artifacts in the Nonlinear IHS in (g) are immediately noticeable. The continuity in lines, as well as sharp contrast changes across regions of the pinkish roofs of an industrial complex in the bottom

Table 5.2 : With-reference performance indicators at reduced resolution on Pleiades dataset.

(worst-best)	QAVE (0-1)	Q (0-1)	sCC (0-1)	SAM (inf-0)	ERGAS (inf-0)
BDSD	.692	.673	.792	2.649	3.049
ATWT	.718	.704	.780	2.226	2.669
GSA	.689	.669	.774	2.535	3.177
GLP-REG-FS	.716	.702	.795	2.329	2.815
Nonlinear IHS	.698	.682	.821	1.873	2.597
Semi-blind Convolution	.712	.700	.750	2.276	19.179
PanNet	.885	.882	.911	1.803	1.440
TA-CNN	**.891**	**.888**	**.933**	**1.509**	**1.295**
PanSRGAN	.917	.889	.960	1.759	1.480
PanColorGAN-GMS	**.956**	**.942**	**.981**	**1.362**	**1.039**
PanColorGAN-PAN	.808	.780	.857	2.116	2.222
PanColorGAN+RD-GMS	.949	.930	.976	1.620	1.219
PanColorGAN+RD-PAN	.794	.763	.850	2.351	2.447

center parts of the image, is preserved only in a few methods. Among those, PanColorGAN-PAN (m) reproduced those features most successfully, followed by BDSD (d), PanColorGAN+RD-PAN (o). Similarly, the spectral or the color reproduction in the results can be gauged from the orange rooftops. Those colors are preserved best in all PanColorGAN models, and PanSRGAN to a degree, whereas the traditional methods all lack the color saturation level of the original multispectral image (a). PanNet (j) and TA-CNN (i) also provided similar visual results to PanSRGAN (k), however, it can be observed that they could not preserve spatial details. The blurring characteristics of the methods are clearly visible, starting with Nonlinear IHS, ATWT, and relatively in all traditional methods except BDSD. Among CNN-based approaches, PanNet, TA-CNN, PanSRGAN, PanColorGAN-GMS, and PanColorGAN+RD-GMS methods show blurrier characteristics with respect to the PanColorGAN-PAN and PanColorGAN+RD-PAN methods, which both clearly outperform all the methods in visual inspection in terms of both structural and spatial properties while keeping spectral properties in an acceptable level when compared to the original multispectral image visually. Although, we obtain higher quantitative scores for PanColorGAN-GMS when compared to the PanColorGAN-PAN variants, it is well-known that higher quantitative scores do not necessarily indicate better perceptual results, as this was also reported in the literature [6].

Table 5.3 : No-reference performance indicators at full resolution on Pleiades dataset.

(worst-best)	D_λ (inf-0)	D_s (inf-0)	QNR (0-1)
BDSD	**.037**	.094	.872
ATWT	.101	.178	.740
GSA	.132	.313	.598
GLP-REG-FS	.089	.150	.774
Nonlinear IHS	.046	.080	.876
Semi-blind Convolution	.123	.227	.678
PanNet	.060	.044	.895
TA-CNN	.041	**.037**	**.920**
PanSRGAN	**.015**	.117	**.869**
PanColorGAN	.042	**.099**	.862
PanColorGAN+RD	.048	.134	.824

5.4 Evaluation of Full Resolution Results

We evaluate the quantitative and qualitative results of the full-resolution experiments in this section.

5.4.1 Quantitative Analysis of Full Resolution Results

Table 5.3 refers to calculated performance measures that require no-reference, as a ground truth or reference pansharpened image does not exist in the real-life full-resolution scenario. TA-CNN provides the best quantitative performance among previous methods followed by PanNet, Nonlinear IHS, and BDSD, whereas both PanColorGAN and PanSRGAN achieve similar results. The measure D_λ focuses on spectral characteristics and D_s focuses on spatial details, whereas QNR is a combination of both measures. In spectral measures, PanSRGAN achieves a good performance in D_λ, whereas TA-CNN achieves the best performance in D_s.

5.4.2 Full Resolution Scenario Visual Results

Figure 5.3 shows full resolution results from all algorithms on the Pleiades test dataset. Images in (a) and (b) refer to the input, i.e. the original panchromatic and multispectral images, respectively. Images in (c)-(h) refer to results produced by traditional methods, whereas (i)-(m) refer to CNN-based methods. Artifacts in results of BDSD (c), (GLP-REG-FS (f), and Nonlinear IHS (g) from traditional methods, as well as in results of PanSRGAN (k), PanColorGAN (l) are apparently visible. Although PanNet

(i) and TA-CNN (j) gave decent results in no-reference metrics, visual results do not support those numbers. They produce more blurry results when they are compared with PanColorGAN+RD. Among the traditional methods, GSA (e) and Semi-blind Convolution (h) produce better results than the former, whereas PanColorGAN+RD (m) provides the best performance. For instance, when the bending corner segments of the white complex structures in the middle of the image are compared, better preservation of continuity of borders is observed in the PanColorGAN+RD method and traditional methods: GSA and Semi-blind Convolution. The sharp edges and high contrast between the white structures and its surroundings is best captured in PanColorGAN+RD and GSA, where the smearing across regions is minimal. In the green fields with tree clusters and vegetation towards top right and bottom left of the scene in the figure, GSA and Semi-blind Convolution preserve the original pattern better than all other methods. One can also observe that because of the low resolution of the MS in (b), the terrain color looks yellow due to the relatively blurry characteristic of the image, whereas the proposed PanColorGAN+RD (m) produces a gray-yellow tone, which matches the colors in other methods. It can be fairly said that all CNN-based techniques are losing the vertical lines of the trees to a degree. This is one limitation we observed in most of the MRA pansharpening methods, including CNN-based methods. In terms of spectral color features, almost all of the techniques including PanColorGANs are observed to capture the original color distributions of the multispectral input image in (b). In terms of spatial features, PanColorGAN+RD shows the best performance, as it includes randomness introduced in its downscaling ratios that increases its robustness to minute resolution variations between the reduced panchromatic and multispectral images.

5.5 Discussions and Transferability

Next, we discuss the transferability capability of the PanColorGAN models, as well as all other baseline methods. For that purpose, the trained CNN-based models on the Pleiades dataset are directly tested on the Digital Globe data in order to assess the transferability of the methods. In Table 5.4, with-reference performance measures for the Digital Globe dataset are given. Again, as in Table 5.2, PanColorGAN-GMS outperformed all other techniques, including traditional methods. PanNet and

Table 5.4 : With-reference performance indicators at reduced resolution on Digital Globe dataset.

(worst-best)	QAVE (0-1)	Q (0-1)	sCC (0-1)	SAM (inf-0)	ERGAS (inf-0)
BDSD	**.832**	.831	**.833**	7.259	4.803
ATWT	.830	**.843**	.827	**6.110**	**4.628**
GSA	.814	.834	.801	7.076	4.952
GLP-REG-FS	.820	.834	.807	6.798	4.777
Nonlinear IHS	.755	.754	.766	6.229	5.808
Semi-blind Convolution	.832	.836	.813	6.062	12.219
PanNet	.690	.681	.633	7.382	6.998
TA-CNN	.673	.665	.622	7.590	7.166
PanSRGAN	.764	.727	.792	7.785	7.430
PanColorGAN-GMS	**.884**	**.845**	**.936**	**6.783**	**4.707**
PanColorGAN-PAN	.835	.796	.879	9.095	6.789
PanColorGAN+RD-GMS	.863	.828	.930	7.746	5.131
PanColorGAN+RD-PAN	.813	.776	.857	9.319	7.182

Table 5.5 : No-reference performance indicators at full resolution on Digital Globe dataset.

(worst-best)	D_λ (inf-0)	D_s (inf-0)	QNR (0-1)
BDSD	.057	.061	.886
ATWT	.091	.146	.777
GSA	.078	.160	.775
GLP-REG-FS	.084	.141	.788
Nonlinear IHS	**.036**	**.046**	**.919**
Semi-blind Convolution	.089	.131	.792
PanNet	.041	.051	.909
TA-CNN	.062	.067	.874
PanSRGAN	**.027**	**.043**	**.930**
PanColorGAN	.040	.073	.890
PanColorGAN+RD	.061	.070	.874

TA-CNN trained on Pleiades Dataset could not provide satisfactory results when they are tested with Digital Globe Dataset which involves different sensor settings. Due to different spatial and spectral resolution characteristics of Pleiades and Digital Globe datasets, a slight decrease in all the quantitative measures are naturally observed for CNN-based methods. Yet PanColorGAN models maintain a slighter decrease when they are compared to other CNN-based methods which are PanNet, TA-CNN, and PanSRGAN.

Table 5.5 refers to no-reference performance measures in the full-resolution mode. Nonlinear IHS achieves the best scores among traditional methods, and PanSRGAN gets the highest scores for the three measures. The real-life pansharpening application with the full-resolution generation deserves further discussions. It is interesting to note that although Nonlinear IHS gives the best quantitative performance with no-reference measures among traditional methods (Table 5.5), it was clearly observed that it performed almost the worst in visual inspection in Figure 5.3. This experiment highlighted the unreliability and mismatch of the no-reference measures against human visual perception. This finding was also reported by Vivone et al. where many pansharpening algorithms are compared [25]. Therefore, in the full-resolution mode, a more reliable evaluation is carried out by visual inspection rather than no-reference quantitative scores.

Figure 5.4 shows visual results from all algorithms in reduced resolution mode on the Digital Globe dataset. Images in (b)-(h) belong to the results of traditional approaches, and (i)-(o) depict results of the CNN-based methods. This is a heterogeneous image patch with many fine man-made structures and fine textural details. Therefore, the artifacts that were observed with Nonlinear IHS (g) before in Figure 5.2 is not that apparent to the eye. However, the first observation that can be easily made is that results of ATWT (c), GLP-REG-FS (f), Nonlinear IHS (g), PanNet (k), TA-CNN (i) and PanSRGAN (k) present blurrier characteristics than the others. Although we were expecting similar results to PanSRGAN, PanNet and TA-CNN gave slightly worse results in terms of spatial quality in reduced resolution tests. As before, the PanColorGAN models are among the best performers, as can be observed over the fine structures in the zoomed flipped C shaped white building. On the other hand, as expected GMS versions of the PanColorGAN provide similar results as the multispectral image while PAN versions preserve spatial details of the reduced panchromatic image. In terms of restoring the color properties, BDSD in (d) and PanColorGAN models (l-o) provides the best visual performance.

Figure 5.5 shows visual results from all algorithms in full resolution mode on the Digital Globe dataset. Images in (a) and (b) refer to the input, i.e. the original panchromatic and multispectral images, respectively. Images in (c)-(h) refer to results produced by traditional methods, whereas (i)-(m) refer to CNN-based methods. The lack of preservation for the spectral and spatial properties of the input panchromatic and multispectral images as well as artifacts are clearly visible in BDSD (c), ATWT (d), GLP-REG-FS (f), Nonlinear IHS (g), semi-blind Convolution (h), TA-CNN (i), PanNet (j) and PanSRGAN in (k). We observe that for the Digital Globe dataset, although the problem of spatial detail disagreement between reduced panchromatic and original multispectral images still persists, it is a less pronounced issue compared to the Pleiades dataset, and this is reflected in the closer quantitative performance results between the PanColorGAN and PanColorGAN+RD. However, when full resolution results in Figures 5.3 and 5.5 are visually inspected, the differences between PanColorGAN and PanColorGAN+RD are clearly observed, where PanColorGAN+RD shows sharper edges and higher contrast than PanColorGAN, which clearly demonstrate the effectiveness of random downsampling in better preservation of spatial details.

A limitation in the development of pansharpening methods is the lack of common datasets. Although standard CNN-based methods, including GAN models, were employed recently for pansharpening, none of those can be evaluated on common data distributions. Naturally, those CNN-based methods were trained and tested on different data distributions, which certainly affects the performance of the models independently from architectural developments. However, our methodological development lies mainly in introduction of a new framework rather than architectural changes, that is why we build a baseline model PanSRGAN with the standard CNN/GAN-based framework, which was crucial to present our improvements in the results.

Our experimental results demonstrate that commonly utilized quantitative image evaluation measures do not necessarily match the expected visual evaluation outcomes. This is not a novel finding, which is also not limited to the domain of satellite imaging. Generally, devising new quantitative image evaluation measures that are faithful to human perceptual evaluations is an open research problem in image analysis.

To summarize our findings, PanColorGAN models are observed to perform at the top among all methods in preserving structural and spatial features of images while keeping the spectral distortion at an acceptable level. This can be asserted for both reduced-resolution and full-resolution modes. In addition, although Digital Globe and Pleiades datasets have different characteristics, PanColorGAN demonstrated better transferability properties than other CNN-based models, as evidenced both quantitatively and qualitatively in our experiments.

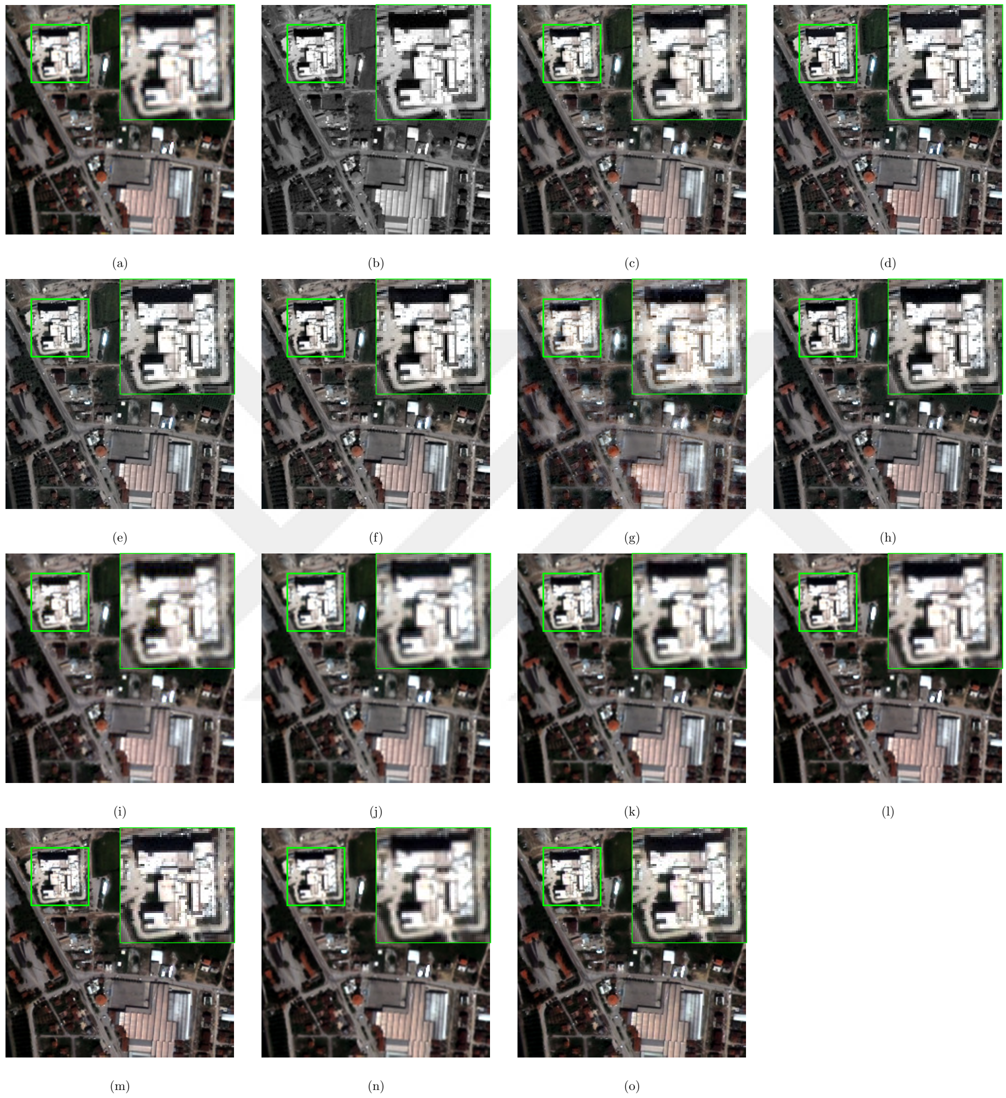

Figure 5.2 : Reduced resolution test results for baseline methods and PanColorGAN models over Pleiades dataset: (a) Multispectral (b) Reduced Resolution Panchromatic (c) ATWT (d) BDSD (e) GSA (f) GLP-REG-FS (g) Nonlinear IHS (h) Semi-blind Convolution (i) TA-CNN (j) PanNet (k) PanSRGAN (l) PanColorGAN-GMS (m) PanColorGAN-PAN (n) PanColorGAN+RD-GMS (o) PanColorGAN+RD-PAN. Region in green box in each picture is zoomed and pasted on the top right for visualization.

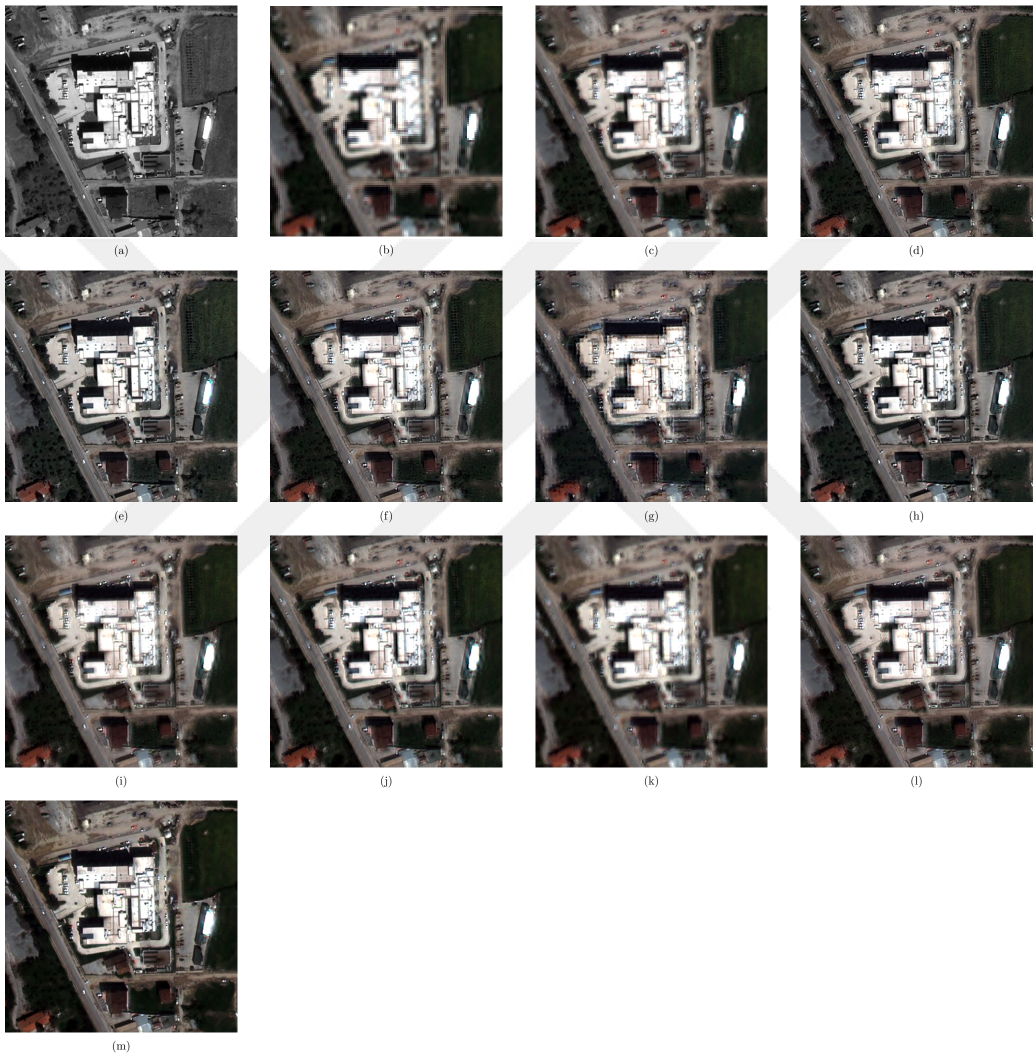

Figure 5.3 : Full resolution test results for Pleiades dataset: (a) Panchromatic (b) Multispectral (c) BDSD (d) ATWT (e) GSA (f) GLP-REG-FS (g) Nonlinear IHS (h) Semi-blind Convolution (i) TA-CNN (j) PanNet (k) PanSRGAN (l) PanColorGAN (m) PanColorGAN+RD.

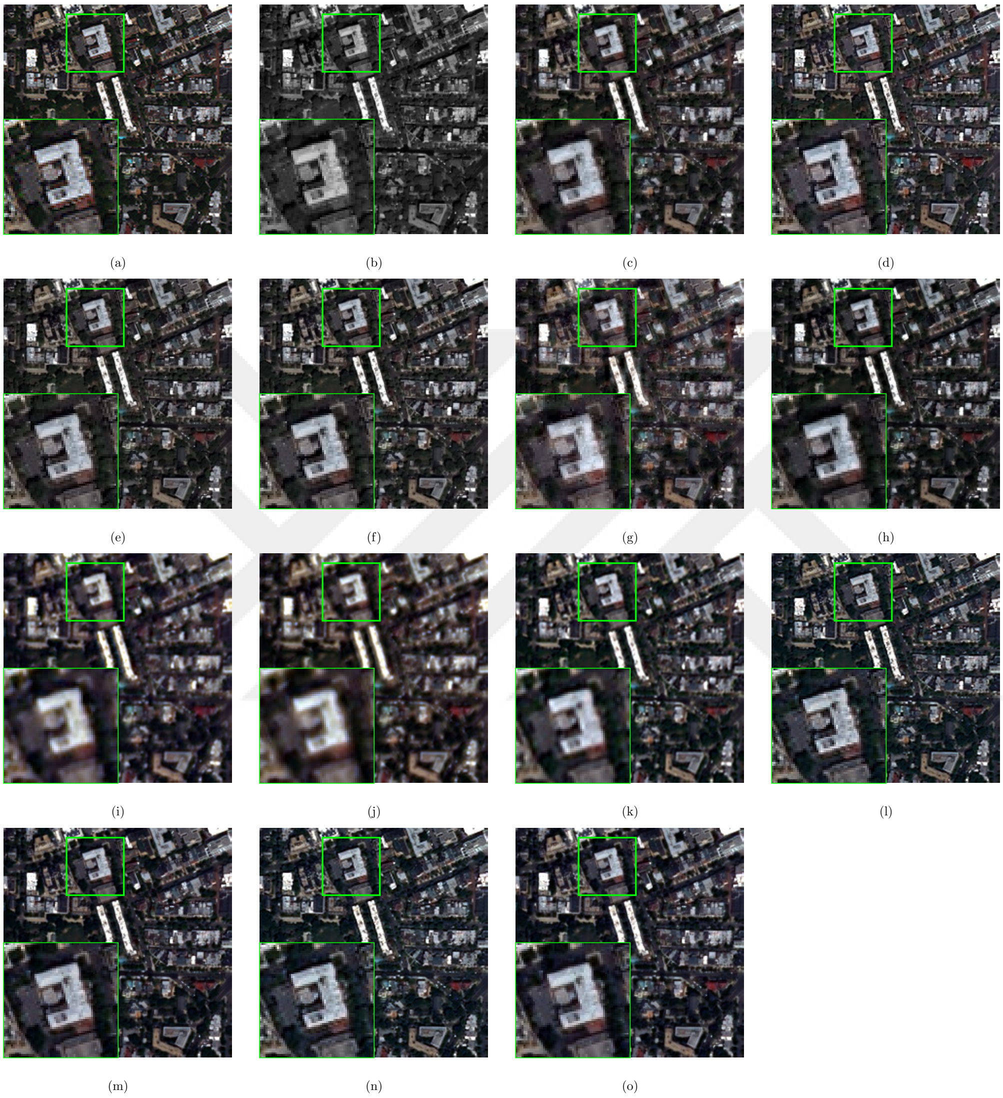

Figure 5.4 : Reduced resolution test results for baseline methods and PanColorGAN models for Digital Globe dataset: (a) Multispectral (b) Reduced Resolution Panchromatic (c) ATWT (d) BDSD (e) GSA (f) GLP-REG-FS (g) Nonlinear IHS (h) Semi-blind Convolution (i) TA-CNN (j) PanNet (k) PanSRGAN (l) PanColorGAN-GMS (m) PanColorGAN-PAN (n) PanColorGAN+RD-GMS (o) PanColorGAN+RD-PAN. Region in green box in each picture is zoomed and pasted at the bottom for visualization.

Figure 5.5 : Full resolution test results for Digital Globe dataset: (a) Panchromatic (b) Multispectral (c) BDSD (d) ATWT (e) GSA (f) GLP-REG-FS (g) Nonlinear IHS (h) Semi-blind Convolution (i) TA-CNN (j) PanNet (k) PanSRGAN (l) PanColorGAN (m) PanColorGAN+RD.

6. CONCLUSIONS AND FUTURE DIRECTIONS

In this thesis, we presented a novel pancolorization framework based on GANs and a guided colorization task for coloring the gray-transformed multispectral images. The new PanColorGAN model, in contrast to earlier CNN-based models with a super-resolution outlook, demonstrated improved structural preservation and reduced blurring effects of the CNN-based methods. In the PanColorGAN, as in traditional GANs, an additional discriminator network is also built in to provide an adversarial loss, to push the generator model towards a better reproduction quality in the pansharpened images. Furthermore, we presented two new spins on the deep neural architecture side:

(i) the injection of the color features from the multispectral image through a second network head into the spatial features extracted from the grayscale image in the first network head of the PanColorGAN;

(ii) introduction of a randomness within a range of scales in the downsampling and upsampling during the network training procedure.

These new contributions helped us to enrich the representation capability of the model in its expressiveness of the multispectral image space, as well as to preserve the spatial details of the panchromatic image space. Particularly, the random downsampling component created a more robust model in terms of generalization to full-scale pansharpening and transferability to other satellites with different MS/PAN spatial resolution ratio without a re-training requirement.

The PanColorGAN demonstrates the state-of-the-art performance in both the reduced-resolution and full-resolution pansharpening modes especially through visual inspection. We articulate that the new deep learning based methods should elaborate extensively on the full-resolution mode results, as it certainly presents the real challenge, particularly due to a potential incompatibility of the spatial details in reduced panchromatic and multispectral images. In addition, the transferability of the PanColorGAN is demonstrated by testing a PanColorGAN network that is previously trained on Pleiades Dataset on the Digital Globe dataset. PanColorGAN achieves excellent spatial resolution, while the spectral resolution it obtains is open to improvement. Finding ways to preserve the spatial and spectral properties in a balanced manner remains an open future research direction in the problem of pansharpening.

As for future extensions of this work, the integration and performance evaluations of the medium spatial resolution satellite images with higher number of multispectral bands and different MS/PAN ratios such as Landsat 8 OLI are planned. PanColorGAN achieves excellent spatial detail preservation, while the spectral information injection efficiency is open to improvement. Enhancing the process of spectral information in PanColorGAN is another future direction for our work.

Another future extension of our work will be developing a model which has a faster inference time. This development can provide a real-time pansharpening process while acquiring best visual results regarding the spatial and spectral properties. In order to enhance the speed of the model, one can use network pruning techniques and EfficientNet-based architectures in our models.

REFERENCES

[1] **Isola, P.**, **Zhu, J.**, **Zhou, T. and Efros, A.A.** (2017). Image-to-Image Translation with Conditional Adversarial Networks, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.5967–5976.

[2] **Goodfellow, I.**, **Pouget-Abadie, J.**, **Mirza, M.**, **Xu, B.**, **Warde-Farley, D.**, **Ozair, S.**, **Courville, A. and Bengio, Y.**, (2014). Generative Adversarial Nets, **Z. Ghahramani**, **M. Welling**, **C. Cortes**, **N.D. Lawrence and K.Q. Weinberger**, editors, Advances in Neural Information Processing Systems 27, Curran Associates, Inc., pp.2672–2680, `http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf`.

[3] **Ledig, C.**, **Theis, L.**, **Huszár, F.**, **Caballero, J.**, **Cunningham, A.**, **Acosta, A.**, **Aitken, A.**, **Tejani, A.**, **Totz, J.**, **Wang, Z. and Shi, W.** (2017). Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.105–114.

[4] **Nazeri, K.**, **Ng, E. and Ebrahimi, M.** (2018). Image Colorization Using Generative Adversarial Networks, ***F.J. Perales and J. Kittler**, editors, Articulated Motion and Deformable Objects*, Springer International Publishing, Cham, pp.85–94.

[5] **Zhu, X.X.**, **Tuia, D.**, **Mou, L.**, **Xia, G.**, **Zhang, L.**, **Xu, F. and Fraundorfer, F.** (2017). Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources, *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36.

[6] **Masi, G.**, **Cozzolino, D.**, **Verdoliva, L. and Scarpa, G.** (2016). Pansharpening by Convolutional Neural Networks, *Remote Sensing*, *8*(7), `https://www.mdpi.com/2072-4292/8/7/594`.

[7] **Yang, J.**, **Fu, X.**, **Hu, Y.**, **Huang, Y.**, **Ding, X. and Paisley, J.** (2017). PanNet: A Deep Network Architecture for Pan-Sharpening, *2017 IEEE International Conference on Computer Vision (ICCV)*, pp.1753–1761.

[8] **Huang, W.**, **Xiao, L.**, **Wei, Z.**, **Liu, H. and Tang, S.** (2015). A New Pan-Sharpening Method With Deep Neural Networks, *IEEE Geoscience and Remote Sensing Letters*, *12*(5), 1037–1041.

[9] **Azarang, A. and Ghassemian, H.** (2017). A new pansharpening method using multi resolution analysis framework and deep neural networks, *2017 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA)*, pp.1–6.

[10] **Liu, X.**, **Liu, Q. and Wang, Y.** (2020). Remote sensing image fusion based on two-stream fusion network, *Information Fusion*, *55*, 1 – 15, `http://www.sciencedirect.com/science/article/pii/S1566253517308060`.

[11] **Liu, X.**, **Wang, Y. and Liu, Q.** (2018). Psgan: A Generative Adversarial Network for Remote Sensing Image Pan-Sharpening, *Proceedings - International Conference on Image Processing, ICIP*, 873–877, `http://arxiv.org/abs/1805.03371,1805.03371`.

[12] **Scarpa, G.**, **Vitale, S. and Cozzolino, D.** (2018). Target-Adaptive CNN-Based Pansharpening, *IEEE Transactions on Geoscience and Remote Sensing*, *56*(9), 5443–5457.

[13] **Wei, Y. and Yuan, Q.** (2017). Deep residual learning for remote sensed imagery pansharpening, *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)*, pp.1–4.

[14] **Wei, Y.**, **Yuan, Q.**, **Shen, H. and Zhang, L.** (2017). Boosting the Accuracy of Multispectral Image Pansharpening by Learning a Deep Residual Network, *IEEE Geoscience and Remote Sensing Letters*, *14*(10), 1795–1799.

[15] **Wald, L.**, **Ranchin, T. and Mangolini, M.** (1997). Fusion of satellite images of different spatial resolutions: Assessing the quality of resulting images, *Photogrammetric engineering and remote sensing*, *63*(6), 691–699, `https://hal.archives-ouvertes.fr/hal-00365304`.

[16] **Krizhevsky, A.**, **Sutskever, I. and Hinton, G.E.** (2012). Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp.1097–1105.

[17] **Russakovsky, O.**, **Deng, J.**, **Su, H.**, **Krause, J.**, **Satheesh, S.**, **Ma, S.**, **Huang, Z.**, **Karpathy, A.**, **Khosla, A.**, **Bernstein, M.** *et al.* (2015). Imagenet large scale visual recognition challenge, *International journal of computer vision*, *115*(3), 211–252.

[18] **Simonyan, K. and Zisserman, A.** (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556*.

[19] **He, K.**, **Zhang, X.**, **Ren, S. and Sun, J.** (2016). Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.770–778.

[20] **Ronneberger, O.**, **Fischer, P. and Brox, T.** (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation, ***N. Navab*, *J. Hornegger*, *W.M. Wells and A.F. Frangi***, *editors, Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, Springer International Publishing, Cham, pp.234–241.

[21] **Jolicoeur-Martineau, A.** (2019). The relativistic discriminator: a key element missing from standard GAN, *7th International Conference on Learning*

Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019, `https://openreview.net/forum?id=S1erHoR5t7`.

[22] **Radford, A.**, **Metz, L. and Chintala, S.** (2016). Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, `http://arxiv.org/abs/1511.06434`.

[23] **Mirza, M. and Osindero, S.** (2014). Conditional generative adversarial nets, *arXiv preprint arXiv:1411.1784*.

[24] **Zhang, R.**, **Isola, P. and Efros, A.A.** (2016). Colorful Image Colorization, ***B. Leibe**, **J. Matas**, **N. Sebe and M. Welling**, editors, Computer Vision – ECCV 2016*, Springer International Publishing, Cham, pp.649–666.

[25] **Vivone, G.**, **Alparone, L.**, **Chanussot, J.**, **Dalla Mura, M.**, **Garzelli, A.**, **Licciardi, G.A.**, **Restaino, R. and Wald, L.** (2015). A Critical Comparison Among Pansharpening Algorithms, *IEEE Transactions on Geoscience and Remote Sensing*, *53*(5), 2565–2586.

[26] **Shettigara, V.K.** (1992). A generalized component substitution technique for spatial enhancement of multispectral images using a higher resolution data set, *Photogram Eng. Remote Sens*, *58*(5), 561–567.

[27] **Ranchin, T. and Wald, L.** (2000). Fusion of high spatial and spectral resolution images: The ARSIS concept and its implementation, *Photogramm Eng. Remote Sens*, *66*(1), 49–61.

[28] **Vivone, G.**, **Restaino, R.**, **Mura, M.D.**, **Licciardi, G. and Chanussot, J.** (2014). Contrast and Error-Based Fusion Schemes for Multispectral Image Pansharpening, *IEEE Geoscience and Remote Sensing Letters*, *11*(5), 930–934.

[29] **Choi, J.**, **Park, H. and Seo, D.** (2019). Pansharpening Using Guided Filtering to Improve the Spatial Clarity of VHR Satellite Imagery, *Remote Sensing*, *11*(6), `https://www.mdpi.com/2072-4292/11/6/633`.

[30] **Meng, X.**, **Shen, H.**, **Li, H.**, **Zhang, L. and Fu, R.** (2019). Review of the pansharpening methods for remote sensing images based on the idea of meta-analysis: Practical discussion and challenges, *Information Fusion*, *46*, 102–113.

[31] **Aiazzi, B.**, **Alparone, L.**, **Baronti, S.**, **Garzelli, A. and Selva, M.** (2006). MTF-tailored multiscale fusion of high-resolution MS and Pan imagery, *Photogrammetric Engineering & Remote Sensing*, *72*(5), 591–596.

[32] **Aiazzi, B.**, **Alparone, L.**, **Baronti, S.**, **Garzelli, A. and Selva, M.** (2003). An MTF-based spectral distortion minimizing model for pan-sharpening of very high resolution multispectral images of urban areas, *2003 2nd GRSS/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pp.90–94.

[33] **Otazu, X.**, **Gonzalez-Audicana, M.**, **Fors, O. and Nunez, J.** (2005). Introduction of sensor spectral response into image fusion methods. Application to wavelet-based methods, *IEEE Transactions on Geoscience and Remote Sensing*, *43*(10), 2376–2385.

[34] **Vitale, S. and Scarpa, G.** (2020). A Detail-Preserving Cross-Scale Learning Strategy for CNN-Based Pansharpening, *Remote Sensing*, *12*(3), `https://www.mdpi.com/2072-4292/12/3/348`.

[35] **Dosovitskiy, A. and Brox, T.** (2016). Generating Images with Perceptual Similarity Metrics Based on Deep Networks, *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, Curran Associates Inc., Red Hook, NY, USA, p.658–666.

[36] DigitalGlobe, `https://www.digitalglobe.com/samples`, date retrieved 18.06.2019.

[37] Open Remote Sensing, `http://openremotesensing.net/`, date retrieved 18.06.2019.

[38] **Garzelli, A.**, **Nencini, F. and Capobianco, L.** (2008). Optimal MMSE Pan Sharpening of Very High Resolution Multispectral Images, *IEEE Transactions on Geoscience and Remote Sensing*, *46*(1), 228–236.

[39] **Aiazzi, B.**, **Baronti, S. and Selva, M.** (2007). Improving Component Substitution Pansharpening Through Multivariate Regression of MS +Pan Data, *IEEE Transactions on Geoscience and Remote Sensing*, *45*(10), 3230–3239.

[40] **Vivone, G.**, **Restaino, R. and Chanussot, J.** (2018). Full Scale Regression-Based Injection Coefficients for Panchromatic Sharpening, *IEEE Transactions on Image Processing*, *27*(7), 3418–3431.

[41] **Ghahremani, M. and Ghassemian, H.** (2016). Nonlinear IHS: A Promising Method for Pan-Sharpening, *IEEE Geoscience and Remote Sensing Letters*, *13*(11), 1606–1610.

[42] **Vivone, G.**, **Simões, M.**, **Dalla Mura, M.**, **Restaino, R.**, **Bioucas-Dias, J.M.**, **Licciardi, G.A. and Chanussot, J.** (2015). Pansharpening Based on Semiblind Deconvolution, *IEEE Transactions on Geoscience and Remote Sensing*, *53*(4), 1997–2010.

[43] **Vitale, S.**, (2019), sergiovitale/pansharpening-cnn-python-version, `https://github.com/sergiovitale/pansharpening-cnn-python-version`, date retrieved 12.01.2020.

[44] PanNet: A deep network architecture for pan-sharpening, `https://xueyangfu.github.io/projects/iccv2017.html`, date retrieved 12.01.2020.

[45] **Zhou Wang and Bovik, A.C.** (2002). A universal image quality index, *IEEE Signal Processing Letters*, *9*(3), 81–84.

[46] **Yuhas, R.**, **Goetz, A.F.H. and Boardman, J.W.** (1992). Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm, *Summaries of the Third Annual JPL Airborne Geoscience Workshop, JPL Publ. 92–14, Vol. 1*.

[47] **Wald, L.** (2000). Quality of high resolution synthesised images: Is there a simple criterion ?, ***T. Ranchin and L. Wald**, editors, Third conference "Fusion of Earth data: merging point measurements, raster maps and remotely sensed images"*, SEE/URISCA, Sophia Antipolis, France, pp.99–103, https://hal.archives-ouvertes.fr/hal-00395027.

[48] **Zhou, J.**, **Civco, D.L. and Silander, J.A.** (1998). A wavelet transform method to merge Landsat TM and SPOT panchromatic data, *International Journal of Remote Sensing*, *19*(4), 743–757, https://doi.org/10.1080/014311698215973, https://doi.org/10.1080/014311698215973.

[49] **Alparone, L.**, **Baronti, S.**, **Garzelli, A. and Nencini, F.** (2004). A global quality measurement of pan-sharpened multispectral imagery, *IEEE Geoscience and Remote Sensing Letters*, *1*(4), 313–317.

[50] **Alparone, L.**, **Aiazzi, B.**, **Baronti, S.**, **Garzelli, A.**, **Nencini, F. and Selva, M.** (2008). Multispectral and panchromatic data fusion assessment without reference, *Photogrammetric Engineering & Remote Sensing*, *74*(2), 193–200.

CURRICULUM VITAE

Name Surname: Furkan Özçelik

Place and Date of Birth: Istanbul, Turkey - 10.10.1996

E-Mail: ozcelikfu@itu.edu.tr

EDUCATION:

- **B.Sc.:** 2018, Istanbul Technical University, Faculty of Computer and Informatics, Department of Computer Engineering
- **M.Sc.:** 2020, Istanbul Technical University, Graduate School of Science Engineering and Technology, Department of Computer Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2018-2020 Turkcell-ITU Graduate Research Fund

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- Ozcelik, F., Alganci, U., Sertel, E. and Unal, G. 2020. Rethinking CNN-Based Pansharpening: Guided Colorization of Panchromatic Images Via GANs, *IEEE Transactions on Geoscience and Remote Sensing,* (Accepted)

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- Trapl, D., Horvaćanin, I., Mareška, V., Ozcelik, F., Spiwok, V. and Unal, G. 2019. Anncolvar: Approximation of Complex Collective Variables by Artificial Neural Networks for Analysis and Biasing of Molecular Simulations, *Frontiers in Molecular Biosciences,* 6(25)