BLIND HYPERSPECTRAL IMAGE RESTORATION FOR UNMANNED AERIAL VEHICLE APPLICATIONS

Victor Sineglazov [1] and Kyrylo Lesohorskyi [2]

[1] Department of Aeronavigation, Electronics and Telecommunication, Kyiv Aviation Institute, Kyiv, Ukraine
[2] Department of Artificial Intelligence, IASA , National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine

ABSTRACT

This work is devoted to the modification of existing blind image restoration algorithms and methodologies for noise and blur elimination in videos and images captured by unmanned aerial vehicles. This work improves on the existing algorithms and methodologies to address the challenges and limitations of existing tools when applied to high-dimensional hyperspectral data by applying channel compression based on 3d convolutions as a dimensionality reduction method. The methods and algorithms described in this paper can be applied in near-real-time and batch-processing scenarios.

A detailed analysis of noise and blur types and their respective sources is provided. An overview of existing methods is given, and their limitations when applied to hyperspectral data are analyzed. A two-stage image restoration approach for hyperspectral data based on is introduced. Proposed algorithms solve the key limitations of hyperspectral data image restoration, providing quality and performance, comparable to non-hyperspectral image restoration.

KEYWORDS

Hyperspectral Imagery, Image Restoration, Recurrent Neural Networks, Unmanned Aerial Vehicles

1. INTRODUCTION

Hyperspectral imagery (HSI) is a set of techniques for analyzing complex objects and surfaces by detecting spectral information from their reflections in hundreds of narrow bands. Usually, hyperspectral cameras operate in both visible and non-visible spectrums, with a typical camera being able to capture 100-500 spectral bands.

Such data has great potential for object classification, monitoring environmental changes, and detecting features invisible to the human eye. However, processing such images has two primary challenges namely high dimensionality and spectral-spatial dependencies.

HSI is widely applied in various industries, including remote sensing of the Earth, environmental monitoring, agriculture, geological research, medical diagnostics, as well as the military and

David C. Wyld et al. (Eds): ADCO, CMCA, NeCoM, NLPI, CSITEC, BDBS, DMML, SP – 2025
pp. 253-268, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.150520

space sectors. For remote sensing applications, HSI cameras are usually mounted on an airborne platforms to maximize the throughput and coverage of hyperspectral data capture process. This, in turn, vastly increases the volume of hyperspectral data that needs to be processed. With the advent of unmanned aerial vehicles (UAVs), they became the primary platform for mounting HIS equipment, as they are able to provide low-altitude capture of data with the pre-programmed flight pass, making them ideal choice for when high resolution data capture is necessary.
Blur and noise removal is tone of the most crucial steps of data proprecessing, especially for data captured by airborne platforms [1]. This step highly increases the effectiveness of intelligent tasks, such as classification, segmentation and subsequent decision making. For small and light UAVs this step is even more important, as their size makes them a less stable platform, making them more susceptible to image degradation caused by external forces.

The problem of noise and blur removal falls under the umbrella of image restoration problems. The area is well-researched, with a variety of methods and algorithms being used to remove unwanted noise and blur from the images. The majority of existing methods are designed with three dimensional data in mind. Applying them to hyperspectral data often proves challenging due to the issues that arise when the number of channels is increased.

Removing noise and blur becomes even more important when it comes to high-risk hazardous applications of hyperspectral imagery, such as landmine removal. Extra caution should be taken to ensure that no artifacts were introduced during the restoration process, as this can lead to type II (or in worst cases type I) errors, slowing down the process significantly or creating additional danger to the personnel by failing to properly classify hazardous object. As such, developing new methods of image restoration for hyperspectral data is extremely important.

2. LITERATURE OVERVIEW

In this section, an overview of specifics and challenges of hyperspectral imagery, types of noise and blur, and existing methods of image restoration is provided.

2.1. Hyperspectral images

Hyperspectral images (HSI) are three-dimensional datasets that contain detailed spectral information for each pixel in the image. Unlike conventional images consisting of three channels (RGB – red, green, blue), hyperspectral images cover a much wider range of wavelengths, divided into hundreds of narrow spectral channels.

As a result, hyperspectral data allows you to obtain unique information about objects that is not available when using only traditional visualization methods.

A hyperspectral image is a three-dimensional tensor (often referenced as hypercube). First two dimensions correspond to spatial representation of the data (coordinates of pixels in the image). Third dimension represents the spectral characteristics of data and contains the values of light intensity in different wavelength ranges (also referred to as spectral bands).

Each pixel in a hyperspectral image captures a vector of object's reflection in different spectral bands, with each value in this vector corresponding to a specific channel. This allows for a detailed spectral signature of an object, material or surface to be created. These spectral signatures are unique to different objects or materials, making it possible to identify and classify them. Hyperspectral images typically cover the visible (VIS), near-infrared (NIR), and mid-infrared (SWIR) ranges of radiation. The spectral channels are arranged with very small

bandwidths (typically 5 to 10 nm), which allows even small differences in light reflectance between objects to be captured. An example hyperspectral image is given in Fig. 1.

The advantages of hyperspectral images are the possibility of obtaining deep information about an object or material. Reflectance depends on physical, chemical and biological properties of the material. This enables the analysis of objects with high accuracy, even if they have the same visual appearance.

Hyperspectral images provide deep analysis of objects and materials due to their unique spectral detail. This makes them one of the most powerful tools for studying and monitoring complex systems and environments. However, it also has several downsides, namely equipment cost, high data dimensionality and spectral-spatial dependency. Due to this, a spectral vector of tens or hundreds of values is formed for each pixel. Despite the high informational density of such data, their processing is difficult due to several challenges.

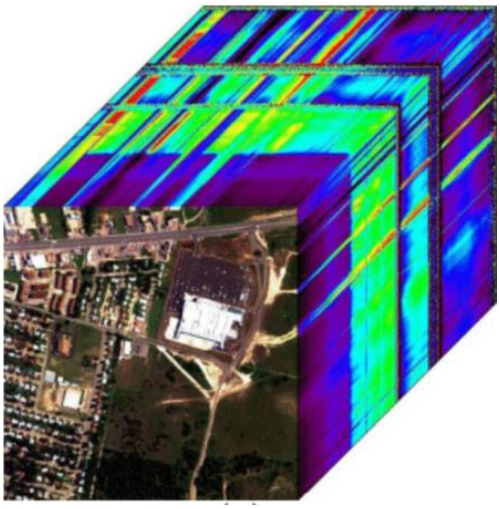

Figure 1. Example of hyperspectral image from Pavia University dataset [2] used in this study

Firstly, the computational complexity increase is polynomial in relation to the number of spectral bands. Every additional spectral band increases the size of spectral dimension by 1, but computational complexity increases by the factor of $W * H$, where W is the width of the image, H is the height of the image. This is especially critical for classification tasks, where algorithms need to analyze each spectral vector and compare it with known classes. The high amount of information complicates the process of training machine learning and deep learning models, making it more resource-intensive.

Secondly, a significant number of spectral features can lead to data redundancy. In hyperspectral images, many spectral channels are interdependent or carry similar information. This creates a problem of data duplication, which reduces the efficiency of analysis and classification algorithms. Common solution to this problem is application of dimensionality reduction techniques. Methods like principal component analysis (PCA), linear discriminant analysis (LDA), or feature filtering are often used to reduce the dimensionality and speed up the processing.

Additionally, spectral complexity can cause "noise" or anomalies in spectral vectors. In real data, variations are observed due to lighting, atmospheric effects, or sensor characteristics. This makes

it difficult to accurately represent objects and requires the use of additional pre-processing methods to normalize the data and filter out noise.

Hyperspectral images are characterized by high dimensionality, which is both their advantage and a significant challenge in processing. Each pixel contains a spectral vector, which usually contains a large number of spectral bands, often more than a hundred. This data structure leads to a significant increase in the amount of information that needs to be processed. The problem of high dimensionality is often described as the "curse of dimensionality". The computational complexity of analysis algorithms grows at exponential rate with regards to the number of spectral channels, which further increases the difficulty of using classical algorithms.

Another side effect of high dimensionality is the need for a larger training dataset for machine learning models in order to prevent over fitting. Limited number of data in high-dimensional setting leads to low generalization ability of models built with machine learning, and negatively impacts overall accuracy.

Another important problem is data redundancy, where many spectral channels carry interdependent or duplicated information. This leads to increased computational costs and poor model performance. Dimensionality reduction techniques enable the separation of useful information from redundant noise, which in turn enables removal of certain dimensions and partially alleviates this problem.

2.2. Noise and blur sources and types

Image restoration tasks are distinguished into blind and non-blind image restoration settings. In the non-blind setting, the type, and possibly parameters, of noise and blur types are known ahead of time. This setting is usually used for a very specific use cases and enable the usage of highly efficient deblurring algorithms. However, in practice, neither noise types nor their parameters are known ahead of time, which corresponds to the blind image restoration problem, which is a setting that will be considered in this work. In blind image restoration problems, however, image restoration is still based on the composition of known noise and blur sources.

2.2.1. Noise

Noise describes an unwanted adversarial alteration of the original signal, which decreases the number of information in the signal. In the context of image or video processing, noise affects pixels of the image, decreasing the amount of information that can be extracted from a given picture or video frame.

In the context of image or video processing, noise is considered an unwanted alteration, that reduces the amount of useful information that can be extracted from a given picture or frame. The majority of noise encountered in images is additive in nature and is usually caused by either manufacturing defects, bit errors in sensor components or environmental factors[3].

Additive white Gaussian noise (AWGN)[4] is one of the most common noise models that can represent a wide range of naturally-occurring noises, such as the thermal vibrations of atoms in conductors, black body radiation or shot noise. This noise is based on the central limit theorem, or rather an assumption that a linear combination of multiple random additive noises will have a Gaussian distribution.

Impulse noise[5] consists happens due to electrical interference or voltage spikes in the transmission medium. It tends to have an approximately flat frequency response over the spectrum range of interest.

Quantization noise[6] is the distortion that occurs when an analog signal is converted into a digital signal with a limited number of discrete amplitudes. Rounding and truncation are typical causes of quantization noise. Quantization is involved to some degree in nearly all digital signal processing, as the process of representing a signal in digital form analogue involves rounding, flooring or ceiling operations during the conversion of continuous signal into discrete values.

Poisson noise[7], also known as shot noise, is the umbrella term for noise that can be modelled as Poisson process. This type of noise is common in electro-optical devices, such as cameras. It is caused by the interference of wave-particle duality with photon counting sensors, which are used to capture the reflectance of objects in the scene.

Speckle noise [8] are granular structures observed in coherent light, resulting from random interference. Speckle patterns happen due to randomization of coherent light, with the most common case being a reflection off an optically rough surface. The speckle effect is a result of the interference of many waves of the same frequency, having different phases and amplitudes, which add together to give a resultant wave whose amplitude, and therefore intensity, varies randomly. It is often the result of atmospheric conditions or manufacturing defect.

While not exhaustive, this list captures the most commonly encountered types of noise. Additive Gaussian noise is one of the most common average noise models in use, and as such is used in this work.

2.2.2. Blur

Blur is an adversarial alteration to the original image due to external factors, such as camera-to-object movement, camera focus, fog, etc. The sharpness of the object's edge decreases and smoothly transitions from one color to another. Blur has a negative impact on the processing of affected regions of the image. There are several common types of blur.

Average blur is spread over the entire surface of the image in both horizontal and vertical directions, and is defined by averaging a circle of radius R as in the following equation:

$$R = \sqrt{g^2 + f^2}, \tag{1}$$

where g, f represents the horizontal and vertical blurring component, and R is the radius of the circular average [5]. This type of blur is commonly used as a synthetic source of blur, however, it is rarely encountered in practice, as blur is rarely uniform in one direction.

Gaussian blur is somewhat similar to AWGN as it is based on the same idea of linear combination of additive blur sources forming normal distribution under the central limit theorem. This type of noise is commonly used for smoothing and can be introduced during the image processing. In this blur, a Gaussian curve is used to calculate a circular gradient form the target kernel location, slowly decreasing in intensity towards the edges. This blur type is defined as a set of parameters for a bell-shaped curve, either in one-dimensional or two-dimensional setting:

$$G(x) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2}{2\sigma^2}}$$

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{\frac{x^2+y^2}{2\sigma^2}}, \quad (2)$$

where *x, y* are centre coordinates of the blur location, σ is the standard deviation of the distribution, which controls the radius and strength of the blur. This blur is a better model for a combination of different blur sources, but requires careful parameter tuning when applied.

Motion blur occurs naturally and is created by a motion between the camera and the target at the time of transition. It is commonly introduced during the image capture and consists of a combination of transitional, rotational, scaling and shift blurs. Due to it's complex nature, no universal analytical representation of this blur exists.

Out-of-focus blurring occurs when a 3-dimensional scene is translated onto 2-dimensional surface during the image capture process. For this phenomenon to occur, the captured object must be located outside of the camera's depth during the exposure stage.

Atmospheric blurring refers to a set of blur types that are caused by natural processes during the image capture. Several factors define the type and strength of atmospheric blur, namely temperature, wind speed, and exposure time

2.3. Image Restoration Methodologies

As discussed in the previous section, image restoration problem is divided into blind and non-blind image restoration. The division into these two subclasses is based on whether the prior for the image degradation operator (blur, noise, or their composition) is known or not. Settings, where the prior is known are considered non-blind image restoration tasks. Non-blind image restoration methods are deterministic and provide faster image restoration in comparison to blind methods. Some of the most famous non-blind image restoration algorithms are the Lucy Richardson algorithm (l-r algorithm) [9], the wiener filter, and a regularized filter [10].

In this work, however, blind image restoration problem [11] is considered due to the complex nature of real-world blur and noise in hyperspectral images. Blind image restoration is more versatile than non-blind approaches, however it usually relies on non-deterministic algorithms that learn low-dimensional embeddings of underlying images to then restore the original image from the said embeddings. The most common way to achiev this is through application of machine learning methods. Neural networks are often chosen for their superior learning capacity in comparison to other machine learning methods.

Convolutional neural networks (CNNs) and their variations are a superior choice [12] for image processing and computer vision tasks, including image restoration problem. They are able to learn embeddings in the convolution layers, which makes their application straightforward. Initially, neural networks were used in the multi-staged approach, where a regression or classification model was used to identify the type of noise and estimate it's parameters to then apply non-blind image restoration techniques. Such approaches are based on less complex network architectures, such as simple multi-layer perceptron [13] or singular valued decomposition [14], however such approaches required extensive fine-tuning for each new application and had poor generalization ability. Further research in this direction considered an approach based on using a convolutional network to learn the prior for deblurring with a half-quadratic optimization is proposed in [15].

More advanced approaches utilize a combination of several neural networks and an external algorithm. One such approach utilizes a classifier to identify the type of the blur kernel and two

sub-networks for parameter identification is proposed by Yan et al. [16]. Alternatively, Sun et al. [17] proposed to use a classifier to describe the linearly non-uniform blur kernel and apply a Markov random field (MRF) to optimize these patch-wise blur kernels.

More recent research focuses on the utilization of more complex network architectures to complete the task of image restoration end-to-end, without relying on any auxiliary algorithms. Several works [18, 19] consider the estimation of the blur kernel and image restoration via a two-stage framework that captures both space and frequency domains. However one challenge these approaches face is that they learn the texture pattern to use as a prior for image restoration, and require a well-defined, consistent texture to work with, limiting it's practical application to domains which meet this requirement, such as text restoration or face restoration.

Unfortunately, the limitation outlined above is considered fundamental and no solution to this problem currently exists. To overcome this limitations, the research is currently focused on using several related images to serve as a prior for a deteriorated image. As such, modern approaches are focused on solving the problem of video frame restoration, rather than just image restoration. While this introduces additional data to serve as a prior for degraded image, video frame restoration also introduces new challenges, namely – temporal dependency, frame alignment and higher data dimensionality, forcing further changes to the architecture.

When it comes to the problem of video frame restoration, there are three methods that are considered state-of-the-art. The first method is called Deblur Network (DBN) [20] and it is based on first pre-procesisng the video with a homography matrix to solve the frame alignment problem, followed by a recurrent neural network to remove the noise and artefacts introduced after th pr-processing stage. Two-shot approach simplifies the workload of the neural network, but pre-processing algorithm itself can introduce noise and artefacts into the video.

As such, more advanced approaches rely on a compelte end-to-end processing. An approach proposed in [21] named Recursive Deblur Network (RDN) is a single-stage approach, and thus does not utilize preprocessing, which makes it inherently faster than DBN. It produces comparable results when the noise is mild, however it performs better when fast motion is present, as the preprocessing step used in DBN tends to add more prominent noise if motion is present in the image.

Recent developments in video deblurring are based on combining existing approaches based on recurrent neural networks and advances in attention mechanisms. The approach proposed by Zhong et al. [22]. The proposed approach improves on existing methods, providing superior restoration capabilities, however, due to the fact that this approach is based on recurrent networks and utilizes a global spatial-attention module, which makes the overall throughput of the network not practical for real-time and near-real-time image reconstruction tasks.

3. PROBLEM STATEMENT

In this section a formal problem statement for HSI restoration is given. Additionally, quality metrics that are used to evaluate the efficiency of image restoration methods are defined.

The problem of blind image restoration is based on the assumption that unknown degradation (also known as distortion) operator is applied to the original image. In this work, a degradation operator is considered as composition of blur and noise:

$$D(x) = N(B(x)), \tag{3}$$

where D is the degradation operator, N is noise operator and B is blur operator, x is an input image. This framework hides the internal details of how noise and blur are actually implemented, however, for better results composition of $N \circ B$ should be non-linear.

The goal of image restoration is to train a restoration learner f that is capable of maximizing the score function (target quality metric) s:

$$argmax_\theta \, s(f(D(x), \theta), x), \tag{4}$$

where θ are the learn weights of the learner f.

Effectiveness of image restoration can be measured objectively or subjectively. Subjective scoring relies on human perception of the image, while objective measure relies on the usage of deterministic metrics to compare original and restored images. In the case of HSI, objective metrics are the best choice, as human eye cannot effectively perceive the multiband data.

To measure the effectiveness of image restoration, two metrics are selected: root mean squared error (RMSE) and peak signal to noise ratio (PSNR).

Root Mean Square Error calculates the difference between the expected value in the training sample and compares it to the value that was produced by the model. It is often used in regression and classification tasks, but can be applied to image restoration as well by using pixel-by-pixel difference to compare original and restored image:

$$RMSE(G, I') = \sqrt{\frac{\sum_{i=1}^{N} (G_i - I_i')^2}{N}}, \tag{5}$$

where N is the batch size, G is the original image, I' - restored degraded image.

PSNR is used to measure the performance of image restoration tasks. It represents a pixel-by-pixel difference between the reference and distorted images' pixels. The PSNR of the image with dimensions m × n is the following:

$$PSNR = 10 * log_{10} \frac{R^2}{\frac{1}{m*n} \sum_{x=0}^{n} \sum_{y=0}^{m} G_{x,y} - I'_{x,y}}, \tag{6}$$

where R is signal strength (maximum possible pixel value in the image), m, n are dimensions of the image, G is the original iamge, I' - restored degraded image, x,y are corresponding pixel positions.

Most commonly, PSNR is positively correlated with image quality, with restoration quality improving as PNSR increases. Optimal PSNR is a perfect match between pixels in, which is equal to the positive infinity [23].

4. METHOD

4.1. Limitations of Existing Methods

One of the primary challenges associated with hyperspectral image processing is the sheer volume of data that needs to be processed. Normal images consist of three channels, while hyperspectral images can have hundreds of bands. This makes processing one hyperspectral

image equivalent to processing ~30 frames of normal images. Considering the fact that several hyperspectral frames are required to serve as a "reference" (or prior) during the deblurring process, this incurs high computational cost and excessive number of parameters for neural network. This is a fundamental limitation of all neural-network based methods, and as such some type of dimensionality reduction has to be applied before the hyperspectral image is successfully processed by neural-network based image restoration algorithms. This leads to two major problems: non-linear local spectral-spatial structures and artefact introduction during the compression / decompression stage.

Non-linear local spectral-spatial structures in the image has two implications for the processing stage. First of all, linear dimensionality reduction algorithms are unlikely to effectively capture the complexity and can lead to the loss of information during the dimensionality reduction. Secondly, the same is true for the noise, and in lesser extent to the blur, in the image. This leads to some spectral bands being more affected by effects of the noise, while others are affected to the lesser extent. As such, compensatory effect should not be equal across all of the channels, as it is likely to lead to image artefacts after the restoration process.
Lastly, it is important to properly model the noise and blur in the image, as the existing datasets for HSI are sparse and usually consist of high quality hand-picked samples, which have good image quality and low frame-per-second, making them hard to use for training image restoration models.

As such, two primary challenges of processing hyperspectral images with existing algorithms are the need for a non-linear dimensionality reduction method and lack of dedicated training dataset. To overcome these issues, in this work a degradation model is introduced to generate synthetic distorted images to solve the dataset scarcity problem with a several dimensionality reduction techniques and modes being analyzed.

4.2. Degradation Model

Degradation model is a one of a key components in solving dataset scarcity. In this work, we follow the general degradation model as outlined in (3) with a linear combination of noise and blur. A combination of blur and noise filters is used to distort the image and create a "low quality" sample.

When building a degradation model, non-linear spectral-spatial structure of the data and noise must be accounted for. As such, two-dimensional noise models introduced in (1) and (2) are poorly fitted to be used as degradation operator. This, in turn, creates a need for a more complex, 3-dimensional noise models.

We also must consider the primary application of the proposed algorithm – processing images captured by UAV during flybys. While proposed algorithms and models are generic in nature, degradation model has to be tailor-fitted for each application. During UAV flybys various types of noise will be encountered, as such AWGN is the most suitable model, as it allows to capture the "average" noise experienced by hypersepctral camera.

This, however, is not true for blur, which is experienced by each channel equally. As such, standard 2-dimensional blur applied to each channel would be sufficient enough to model the blur properly. Considering that during the applications of the HIS, during the flyby UAV will not be focused on any specific object, but rather scan the surface directly beneath it. This makes motion blur and rapid shakes the most prominent source of blur in the image.

As such, in this study we apply a combination of 3-dimensional AWGN as the noise component and either a 2-dimensional Gaussian blur or two-dimensional motion blur as the blur component. The 3-dimensional AWGN is one of the generalizations of N-dimensional normal distribution in the following form:

$$G(x,y,z) = \frac{1}{2\pi^{3/2}\sigma^3} e^{\frac{(x-x')^2+(y-y')^2+(z+z')^2}{2\sigma^2}}, \quad (7)$$

where σ is the deviation parameter of the distribution and controls the "strength" of the noise, *x'*, *y'*, *z'* are the centre coordinates of the noise, *x*, *y*, *z* are the coordinates of the current pixel. The noise is introduced in a random centre *x'*, *y'*, *z'* drawn from a uniform discrete distributions:

$$x' \in [0, W], y' \in [0, H], z' \in [0, C] \quad (8)$$

where *W* is the width of the image, *H* is the height of the image. Applying this filter introduces the noise in the shape of concave quadratic plane (Fig. 2) that is centered somewhere in the image, but only affects a relatively small patch of it.

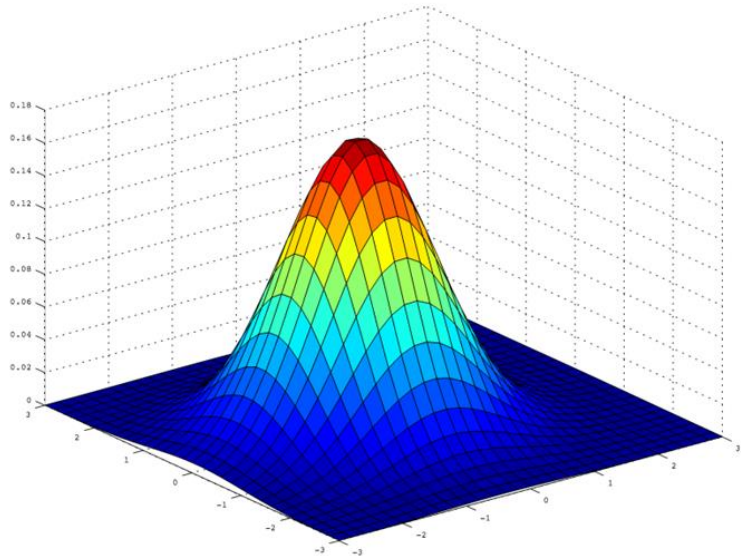

Figure 2. The shape of 3-dimensional Guassian distribution

The strength of the noise represents how strong the blur will be. To achieve a good mix of strong and weak noise. To emulate it, during each instance of noise introduction, the distribution parameter is pulled from a normal distribution with variance parameter set to *0.2 * W*. Please note that in this research we use images where *H* = *W*, which ensures symmetry of the noise across the dimensions. If this ratio is not maintained, however, separate scaling factor should be utilized to ensure that noise is symmetric.

To emulate blur we use two types of blur: Gaussian centre blur and motion blur via convolution kernel. Gaussian blur uses 2-dimensional version as described in (2) and follows the same model to emulate blur strength as the noise component, except the variance of normal distribution is set to just 0.2.

Motion blur is emulated through one-hot encoded convolution kernel filter. The initial kernel is formed by creating a kernel of size *H*/3 × *H*/3 initialized with zeroes. Then, the centre column is initialized with the random value drawn from a normal distribution with $\sigma = 0.5$. This emulates

the strength of motion blur, but also ensures that overall impact of blur is limited. The final step of blur kernel formation is to rotate it to simulate the movement in different direction. To achieve this, an Euclidian rotation is used:

$$Rv = \begin{pmatrix} xcos\theta - ysin\theta \\ xsin\theta + ycos\theta \end{pmatrix}, \tag{9}$$

where θ is the rotation degree, $\theta \in [0, 2\pi]$.

The blur is applied across all channels, simplifying the application of the filters. The scheme outlined above ensures that blur strength can be controlled to create varied augmentations, as can be seen in Fig. 3.

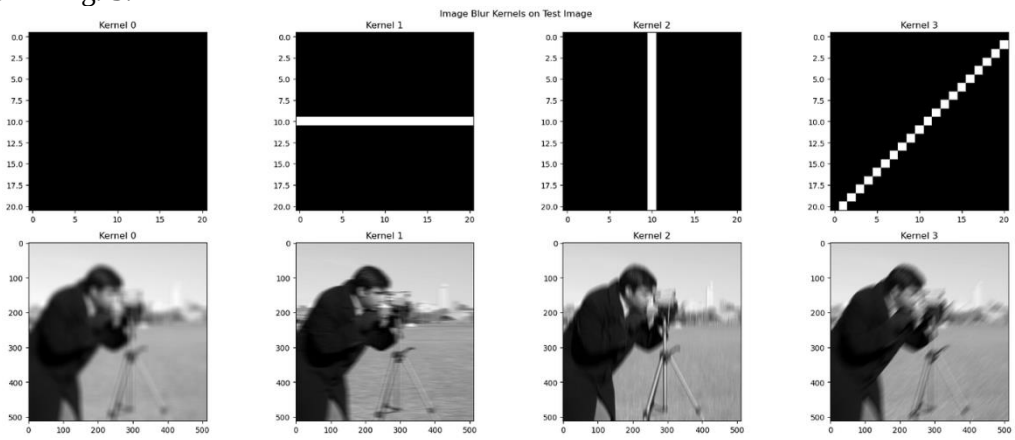

Figure 3. Example of motion blur kernel and application

4.3. Channel Compression

Channel compression is an important step in the model that compacts and encodes the raw bands in input data to ensure that it can be processed by the following neural network. Two possible approaches to the channel compression are dimensionality reduction methods, such as PCA, or end-to-end compression with 3D convolutions.

Dimensionality reduction is the most commonly used technique, as it is well-researched and is easy to apply. Non-linear dimensionality methods, such as Isomap [24] are often used as pre-processing step in HSI processing, however the main drawback of applying this methods is the fact that it might yield different number of "information dense" channels. While not a major drawback, this would require to re-train the underlying neural network's adaptive layer each time a new dataset is used, which in turn hurts complicates fine-tuning of the model greatly.

Alternative approach would be to utilize "learnt" dimensionality reduction. This approach is more generic and utilizes 3D convolution layers in front of the neural network to compress the channels.

In this work, the latter approach is used as image restoration is a type of sequence-to-sequence problem, which requires matching both the original input and output. To remove any bias that might be introduced by dimensionality reduction methods, 3D convolutions are used as channel compression methods. In practice, however, image restoration is considered a pre-processing step, in which scenario using dimensionality reduction channel compression might be beneficial, as next steps are likely to need to reduce dimensionality as well, in which case non-linear

dimensionality reduction method is the recommended way to set up the channel compression. A 5-layer encoder a decoder path with 5×5×10 convolution kernel are used for both constriction and extension pathway.

4.4. Neural Network Architecture

When it comes to the neural network architecture, the approaches can be broken down into two types: one-pass restoration and multi-pass recurrent restoration. In case of the one-pass approach, several neighbouring frames are squashed into a single "multi-channel" frame that has internal temporal coupling. The assumption is made that if these frames are close enough together, the neural network will be able to learn appropriate frame alignment and use squashed channel frames as a prior for the degraded frame restoration. These approaches tend to be more computationally efficient than recurrent algorithms.

In the case of HSI, however, these methods have severe drawbacks. Even after the channel compression, the image is likely to contain multiple (10-15) channels. Merging together multiple frames will result in the same problem as just processing raw hyperspectral data – the number of channels is too high, causing the curse of dimensionality. Additionally, frequency of HSI is usually much lower than that of normal cameras, which means that each frame is further apart from each other, making the assumption of two frames being nearby invalid.
As such, a multi-pass recurrent neural network is a preferable approach in case of HSI restoration. For image restoration, the most commonly used neural network architecture is encoder-decoder. This architecture consists of constriction pathway, that slowly extracts a feature map from the image by slowly reducing dimensionality through convolution operation, followed by the extension pathway that restores the feature vector into the original image.

A traditional U-net [25] architecture was used to build an encoder-decoder core. One notable difference is that our approach utilizes residual recurrent connections from past layers to keep temporal context. The architecture is presented in the Fig 4.

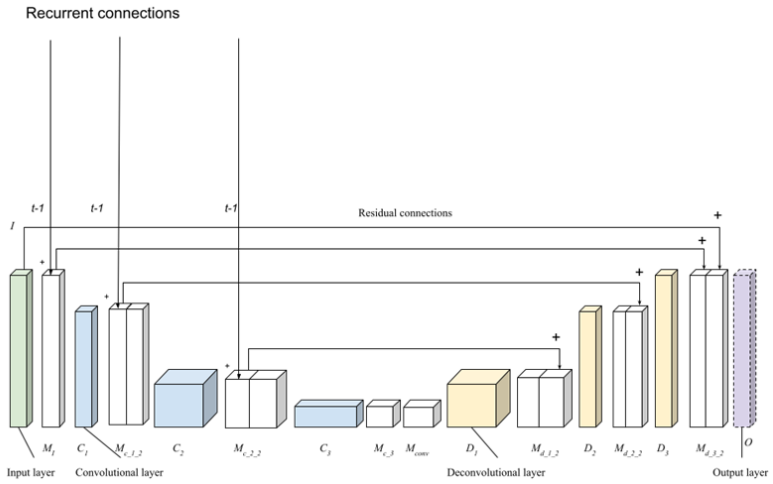

Figure 4. Network architecture (excluding channel compression pathway)

A more detailed layer configuration is outlined in the table 1. In the runtime, 3 frame setup (2 reference, 1 distorted) are used to restore the image as it allows for great balance of quality and performance. The deconvolution pathways are disabled in reference frames to further boost performance.

Table 1. Neural Network Layer configration

Layer	Filter Size	Stride	Output Shape	Activation Function
I			$H \times W \times 5C$	
M_i	$3 \times 3 \times 128$	1	$H \times W \times 128$	ReLU
C_1	$3 \times 3 \times 128$	2	$H/2 \times W/2 \times 128$	
$M_{C\_1\_2}$	$3 \times 3 \times 128$	1	$H/2 \times W/2 \times 128$	ReLU
C_2	$3 \times 3 \times 256$	2	$H/4 \times W/4 \times 256$	
$M_{C\_2\_2}$	$3 \times 3 \times 256$	1	$H/4 \times W/4 \times 256$	ReLU
C_3	$3 \times 3 \times 512$	2	$H/8 \times W/8 \times 512$	
M_{C_3}	$3 \times 3 \times 512$	1	$H/8 \times W/8 \times 512$	ReLU
M_{conv}	$1 \times 1 \times 512$	1	$H/8 \times W/8 \times 512$	ReLU
D_1	$4 \times 4 \times 512$	1/2	$H/4 \times W/4 \times 256$	
M_{D1_2}	$3 \times 3 \times 256$	1	$H/4 \times W/4 \times 256$	ReLU
D_2	$4 \times 4 \times 256$	1/2	$H/2 \times W/2 \times 128$	

4.5. Dataset Generation and Neural Network Training

In this research two datasets – Indian Pines and Pavia University [2] are used. The datasets consist of stitched hypercubes of data, as such additional slicing is necessary to create frames for deblurring. Sliding window is used to "slice" the hypercube into multiple overlapping images, simulating the data capture during the fly-by.

The set of frames is generated by taking three neighbouring slices from the sliced hypercube, chaining them in 3-frame sequences and applying the degradation operator described in 4.2 to the final frame to generate "low-quality" samples. This process is repeated until a sizeable dataset is built for the model learning. In this research datasets of 10 000 samples were used.

The training process consist of two steps: learning data representation and learning image restoration. During the first step the network is trained in the autoencoder mode to learn the internal representation of the hyperspectral images. The first stage uses mean squared error loss function with ADAM optimizer during the learning process. Empty feature vectors are used in palce of recurrent values. Learning rate is set to 0.01, decreasing up to 0.0001 over the training period. The network is trained in batches of 64 images with early stopping or for 5000 epochs.
After the autoencoder is trained, the weights in the channel compression layers are frozen and only the restoration decoder is fine-tuned for image restoration. It uses the same settings as the auto-encoder stage, however, in this case residual connections are activated and the neural network learns the restoration on the sequences of three images.

After the training is concluded, the deconvolution pathways for all recurrent layers but final iteration are disabled to minimize the compute required.

5. RESULTS AND DISCUSSION

The evaluation of the dataset was performed on the Indian Pines and Unviersity of Pavia datasets, mentioned in the previous section.

Both of the datasets are captured by an aircraft-mounted hyperspectral sensor during a fly-by over an urban area. The dataset contains 103 specral bands in the range of 430 to 860 nanometes, which covers both visible and near-infrared spectra. The resolution of the dataset is 1.3 meters per pixel. The datasets were split into a training and validation datasets in a ratio of 80-20%. Considering the limited size of each dataset, proposed model were not able to achieve great generalization ability, and as such is tested independently on each dataset. After the training, validation data is used to evaluate the results. PSNR (3) is used to measure the quality in several settings. The results are presented in Table 1

Table 2. PSNR for the proposed method

Noise Type	**Dataset**	**PSNR**
No noise / blur	Pavia University	23.538
Weak noise / Weak blur	Pavia University	24.281
Average nosie / Average blur	Pavia University	24.351
Strong noise / Strong blur	Pavia University	26.3781
Average noise / No blur	Pavia University	24.3475
No noise / Strong blur	Pavia University	26.287
No noise / blur	Indian Pines	31.258
Weak noise / Weak blur	Indian Pines	33.874
Average nosie / Average blur	Indian Pines	34.561
Strong noise / Strong blur	Indian Pines	37.115
Average noise / No blur	Indian Pines	34.552
No noise / Strong blur	Indian Pines	34.543

These results are comparable to metrics achieved by conventional image restoration methods for non-hyperspectral images on NITRE challenge dataset.
Performance-wise the model was able to achieve approximately 220ms inference time for three frame setup, which corresponds to ~6 FPS. Further performance boost is possible by caching inference result for previous layers during continuous operation, however the model still requires further optimization to be used in the real-time scenario. Near-real-time scenarios, however are still possible with the current level of performance.

One of the key observations is that the model has strong resiliency towards strong noise, but struggles with a strong blur. It can be seen in both cases that noise strength has limited impact on the value of PSNR metric.

On the other hand, strong noise makes severe impact on model quality and is the key contributor to model degradation. Likely explanation of this lies in the fact that we use a localized version of the noise, which has limited impact. Additionally, noise is additive in nature, which makes reversing it's effects on the image easier. Blur, on the other hand, smothers the features across relatively large patches of the image, which makes it much harder to reverse.

Two shortcomings in the proposed approach can be highghlited, namely – small convolutional kernel in the channel compression pathways may lead to loss of information in the image restoration part. The second problem lies in the way degradation strength is modelled. Due to using normal distribution, "strong" blur samples are fairly rare in the overall population. At the same time, they are also the hardest to learn the weights for. Increasing the occurrence of "strong" blur samples in the training population is likely to increase the performance of the restoration model.

6. CONCLUSIONS

The proposed methods shows promising results, although still requires extra tuning to be used in the real-world scenarios. Overall neural network architecture is straightforward and is based on well studied 3D convolutions and encoder-decoder architecture, although some parameters still require fine-tuning. The algorithm can be applied to either real-time UAV feed or to the flight footage recording to both improve the visual fidelity for human perception or improve the accuracy of subsequent processing by other intelligent methods.

One of the biggest limitations of this research is the lack of dedicated, high-quality dataset for hyperspectral image restoration tasks. Datasets used in this research are of high quality, but low resolution (1.3 meters per pixel) makes it impossible to test the performance of the algorithm when on the scenes that contain small and fast objects, which would introduce additional blur. Another limitation of the proposed algorithm is computational intensity of the algorithm due to it's recurrent architecture. This limits the usage of the algorithm to mostly post-processing, as it cannot be applied on the UAV itself due to power consumption concerns.

Future research includes two areas of improvement: decreasing the computational complexity and improving the stability in the presence of the strong noise.

Apart from the dataset, future research will be focused in two strategic directions: improving the performance of image restoration to match the requirements of real-time image processing and improving the stability of the model in the presence of strong noise following the recommendations outlined in the results and discussion section.

REFERENCES

[1] Sineglazov, V., Lesohorskyi, K., & Chumachenko, O. (2024, February). Faster Image Deblurring for Unmanned Aerial Vehicles. In 2024 2nd International Conference on Unmanned Vehicle Systems-Oman (UVS) (pp. 1-6). IEEE.

[2] Huang, Xin, and Liangpei Zhang. A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy. International Journal of Remote Sensing 30.12 (2009): 3205-3221.

[3] Kaur S., "Noise types and various removal techniques," Int. Journal of Adv. Research in Electronics and Com. Eng. (IJARECE), 2015 , Feb, vol. 4, no. 2, pp. 226-30.

[4] P. Bergmans, "A simple converse for broadcast channels with additive white gaussian noise (corresp.)," IEEE Trans. on Information Theory, 1974, vol. 20, no. 2, pp. 279-280.

[5] Shongwe T., Vinck AJ, Ferreira H. C., "A study on impulse noise and its models", SAIEE Africa Research Journal, 2015, vol. 106, no. 3, pp. 119-131.

[6] Fan A., Stock P., Graham B., Grave E., Gribonval R., Jegou H., Joulin A., "Training with quantization noise for extreme model compression," arXiv preprint arXiv:2004.07320. 2020 Apr 15.

[7] Hasinoff S., "Photon, Poisson Noise," Computer Vision, A Reference Guide, 2014, vol. 16, no. 4, pp. 1.

[8] Bianco V., Memmolo P., Leo M., Montresor S., Distante C., Paturzo M., Picart P., Javidi B., Ferraro P., "Strategies for reducing speckle noise in digital holography," Light: Science & Applications, 2018 Aug 1, vol. 7, no. 1, pp. 48.

[9] He Z., Wang H., Li Y., Zhang Z., Zhang Y., Bi H., He Y., "A deconvolutional reconstruction method based on Lucy–Richardson algorithm for joint scanning laser thermography," IEEE Transactions on Instrumentation and Measurement, 2020, Oct. 30, vol. 70, pp. 1-8.

[10] Takeda H., Farsiu S., Milanfar P., "Deblurring using regularized locally adaptive kernel regression," IEEE transactions on image processing, 2008, Mar. 14, vol. 17, no. 4, pp. 550-563.

[11] Lai W., Huang J., Hu Z., Ahuja N., Yang M., “A comparative study for single image blind deblurring,” In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1701-1709.
[12] Zgurovsky, M., Sineglazov, V., Chumachenko, E., Zgurovsky, M., Sineglazov, V., & Chumachenko, E. (2021). Classification and Analysis of Multicriteria Optimization Methods. Artificial Intelligence Systems Based on Hybrid Neural Networks: Theory and Applications, 59-174.
[13] Schuler C., Christopher Burger H., Harmeling S., Scholkopf B., “A machine learning approach for non-blind image deconvolution,” In Proc. of the IEEE conf. on computer vision and pattern recognition, 2013, pp. 1067-1074.
[14] Ma L., Xu L., Zeng T., “Low rank prior and total variation regularization for image deblurring”, Journal of Scientific Computing, 2017, Mar, vol. 70, no. 3, pp. 1336-57.
[15] Zhang K., Luo W., Zhong Y., Ma L., Stenger B., Liu W., Li H., “Deblurring by realistic blurring,” In Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 2737-2746.
[16] Yan R., Shao L., “Blind image blur estimation via deep learning,”, IEEE Transactions on Image Processing, 2016, Feb. 26, vol. 25, no. 4, pp. 1910-1921.
[17] Sun J., Cao W., Xu Z., Ponce J., “Learning a convolutional neural network for non-uniform motion blur removal,” In Proc. of the IEEE conf. on computer vision and pattern recognition, 2015, pp. 769-777.
[18] Chakrabarti A., “A neural approach to blind motion deblurring,” In Proc. of Computer Vision–ECCV 2016: 14th European Conf., Amsterdam, 2016, Oct. 11-14, pp. 221-235.
[19] Schuler C., Hirsch M., Harmeling S., Schölkopf B., “Learning to deblur,” IEEE transactions on pattern analysis and machine intelligence, 2015, Sep. 23, vol. 39, no. 7, pp. 1439-1451.
[20] Zhan, Zongqian, "Video deblurring via motion compensation and adaptive information fusion," Neurocomputing, 2019, vol. 341, pp. 88-98.
[21] Su S., Delbracio M., Wang J., Sapiro G., Heidrich W., Wang O., ”Deep video deblurring for hand-held cameras”. In Proc. of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1279-1288.
[22] Pathak D., Krahenbuhl P., Donahue J., Darrell T., Efros AA., “Context encoders: Feature learning by inpainting,” In Proc. of the IEEE conf. on computer vision and pattern recognition, 2016, pp. 2536-2544.
[23] Chai T., Draxler R., “Root mean square error (RMSE) or mean absolute error (MAE),” Geoscientific model development discussions, 2014, Feb. 28, vol. 7, no. 1, pp. 1525-34.
[24] Lou, J., Jin, T., Zhou, Z.: Feature extraction for landmine detection in uwb sar via swd and isomap. Progress In Electromagnetics Research 138, 157–171 (2013)
[25] Krithika Alias AnbuDevi, M., and K. Suganthi. "Review of semantic segmentation of medical images using modified architectures of UNET." Diagnostics 12.12 (2022): 3064.

AUTHOR

Victor Sineglazov is a Doctor of Engineering Science and a Professor, serving as the Head of the Aviation Computer-Integrated Complexes Department at the Faculty of Air Navigation Electronics and Telecommunications, National Aviation University in Kyiv, Ukraine. He graduated from Kyiv Polytechnic Institute in 1973. His research areas include air navigation, air traffic control, identification of complex systems, wind and solar power plants, and artificial intelligence. He has authored more than 700 publications.