Mapping RGB-to-NIR with Pix2Pix Image-to-Image Translation for Fire Detection Applications

Amila Akagić, Emir Buza
University of Sarajevo
Faculty of Electrical Engineering
Zmaja od Bosne bb, 71000 Sarajevo, BiH
{aakagic, ebuza}@etf.unsa.ba

Marko Horvat
University of Zagreb
Faculty of Electrical Engineering and Computing
Department of Applied Computing
Unska 3, 10000 Zagreb, Croatia
marko.horvat3@fer.hr

Abstract. *In this research paper, we present a novel application of image-to-image translation techniques for fire detection applications. The focus of the study is on translating RGB images into Near InfraRed (NIR) images, which can serve as a preprocessing step for fusion or other methods that benefit from such information. The Pix2Pix model is employed to generate NIR images from existing RGB images, thereby adding an extra data source. The experimental results show the ability of the model to successfully learn the translation process, capturing the desired characteristics in the generated NIR images. The histograms and structural similarity index confirm the fidelity of the translations, with an average index value of 82.42%. The model effectively detects the position and shape of fire in the images, although some details on the edges may appear less prominent. Data augmentation could be employed to enhance the ability of the model to produce high-quality NIR images in various scenarios.*

Keywords. Computer Vision, Image Processing, Deep Learning, Image-to-Image Translation, NIR

1 Introduction

Computer vision-based wildfire prediction and detection systems have gained significant attention and popularity in recent years (Buza and Akagic, 2022; Akagic and Buza, 2022). The advancements in technology, especially in the fields of machine learning and deep learning, have enabled the analysis of large volumes of digital images with greater accuracy and efficiency (Šabanović et al., 2023; Kapo et al., 2023). These systems hold great promise for detecting and predicting wildfires. In this process, unmanned aerial vehicles (UAVs) equipped with computer vision-based technology play a crucial role in monitoring and fighting wildfires.

In recent trends of computer vision and image processing, there is compelling evidence supporting the idea that the inclusion of Near InfraRed (NIR) information is beneficial for pattern recognition, as it offers significant advantages in capturing a comprehensive representation of a scene across various applications (Ghiass et al., 2014; Mishra et al., 2022). The inclusion of NIR information is motivated by its ability to provide valuable insights that enhance accuracy and utility in real-world scenarios. Unlike prevailing methods that predominantly analyze 2D images within the visible light spectrum (Brdjanin et al., 2020; Dardagan et al., 2021), which are vulnerable to variations in environmental illumination that can degrade their performance, NIR bands exhibit resilience to such fluctuations, making them a valuable and reliable data source for image processing.

The NIR data provides a promising avenue for the development of a novel methodology aimed at analyzing multi-modal images. This innovative approach holds the potential for augmenting accuracy in pattern recognition within the domain of image processing. Its efficacy is particularly pronounced in situations where labeled training data is limited or unattainable, thereby addressing the constraints imposed by specific regions or circumstances. Moreover, this approach exhibits versatility across various data sources, encompassing satellite or Unmanned Aerial Vehicle (UAV) imagery, which facilitates the identification of intricate patterns and anomalies without necessitating human labeling or intervention.

In this research paper, we introduce a pioneering application of image-to-image translation techniques targeted at fire detection. Our focus centers on the translation of initial RGB images into Near InfraRed (NIR) images, which can serve as a preprocessing step for fusion or other types of methods that can benefit for such information. Our approach can be used to generate NIR images from existing RGB images, thus creating an additional source of data. The RGB and NIR data sources can then be combined or fused to increase the resilience of the algorithm for the increased benefit of pattern recognition. Such algorithms can leverage several key advantages, including heightened resilience to variations in environmental illumination and improved spatial resolution. These benefits facilitate the pre-

cise identification of areas impacted by fires, thereby enabling more efficient and timely wildfire detection. Such advancements play a pivotal role in safeguarding human life, preserving the environment, and mitigating the deleterious consequences associated with wildfires.

In the context of wildfire detection, RGB-NIR data can be particularly beneficial. It enhances the spatial resolution of the images and allows for the differentiation of vegetation types based on their reflectance patterns in the NIR spectrum. This capability enables a better identification of areas affected by flames versus those that are not. In recent literature, there are very limited data sources for NIR images, as opposed to the RGB fire and/or flame images. Thus, this paper tries to address this limitations and provide valuable insights into generating new data sources. We believe that this approach can be used for other applications as well.

The rest of this work is organized as follows. The related work is provided in Section II. The proposed method is described in section III. In Section IV, the evaluation of the proposed method and results are described. Finally, Section V concludes the paper.

2 Related Work

As previously mentioned, the inclusion of Near InfraRed (NIR) information has been shown to significantly enhance and streamline the performance of various image processing and computer vision tasks. This has been exemplified in several applications documented in the literature. For instance, in studies such as (Rüfenacht et al., 2013) and (Salamati et al., 2011), NIR information was utilized to effectively remove shadows from images. The elimination of shadows in images holds the potential to enhance the accuracy and efficacy of tasks such as tracking, segmentation, and object detection. Shadow boundaries often result in confusion with different surfaces or objects, thus it becomes advantageous to eliminate them entirely from the image.

In (Brooksby et al., 2003), the authors investigate the potential of integrating Magnetic Resonance Imaging (MRI) and Near-Infrared (NIR) imaging modalities to achieve noninvasive, high-resolution maps of optical properties. In (Schaul et al., 2009) and (Feng et al., 2013), two image dehazing approaches that leverage the combination of color and NIR images are proposed. In (Clarke, 2004), the authors explore two methods for extracting process-related information from NIR microscopy data cubes. Additionally, NIR has found applications in image enhancement (Matsui et al., 2011), image registration (Firmenichy et al., 2011), image color restoration (Lv et al., 2022), prediction of quality attributes (Kamruzzaman et al., 2012), and many others.

In the context of image-to-image translation, both NIR-to-RGB and RGB-to-NIR conversions have been explored. The NIR-to-RGB translation has been investigated in (Dou et al., 2019), where it was applied to face image translation. In (Sun et al., 2019), the authors proposed a novel asymmetric cycle GAN for NIR to RGB domain translation, specifically focusing on NIR colorization. While (Yan et al., 2020) introduced a new method for NIR-to-RGB translation, utilizing a U-net based neural network to learn texture information and a CycleGAN based neural network to extract color information. In (Shukla et al., 2022), the authors presented a technique for high-resolution NIR prediction from RGB images, targeting plant phenotyping applications. While both approaches exist, the literature tends to have a relatively greater emphasis on the NIR-to-RGB translation.

The RGB-to-NIR approach has gained attention in agricultural applications for determining crop parameters that are not visible to the human eye, as discussed in (Aslahishahri et al., 2021). The authors investigate image-to-image translation techniques to generate a NIR spectral band solely from a RGB image in aerial crop imagery. Similarly, (Sa et al., 2022) synthesize NIR information from RGB input using a data-driven, unsupervised approach without the need for manual annotations or labels. Their focus is on enhancing the system for fruit detection. (Yuan et al., 2020) explore the use of a conditional Generative Adversarial Network (cGAN) structure to generate a NIR spectral band conditioned on the input RGB image. (Ciprián-Sánchez et al., 2021) provides a quantitative demonstration of the feasibility of applying deep learning-based fusion methods to infrared imagery from wildfires. They introduce a novel artificial IR and fused image generator called FIRe-GAN. The goal of this method is to fuse RGB with NIR information while trying to preserve the consistent color.

The goal of this paper is to find a way to translate an RGB image and generate a corresponding NIR image as output. An inference module is created to generate a domain image-to-image translation, e.g. translation of RGB to NIR images.

2.1 Image-to-image Translation

Image-to-image translation refers to a task of converting an input image from one domain to a different domain image while preserving its essential content. In the case of RGB to NIR image translation, the goal is to transform an RGB image into its corresponding NIR representation. RGB images are composed of three color channels: red, green, and blue, which are visible to the human eye. On the other hand, NIR images capture light in the near-infrared spectrum, which is beyond the range of human vision.

In recent literature, image-to-image translation techniques are often based on deep learning models like generative adversarial networks or variational autoencoders (VAEs) (Liu et al., 2017; Zhu et al., 2017; Pang et al., 2021). For training a model, a dataset that con-

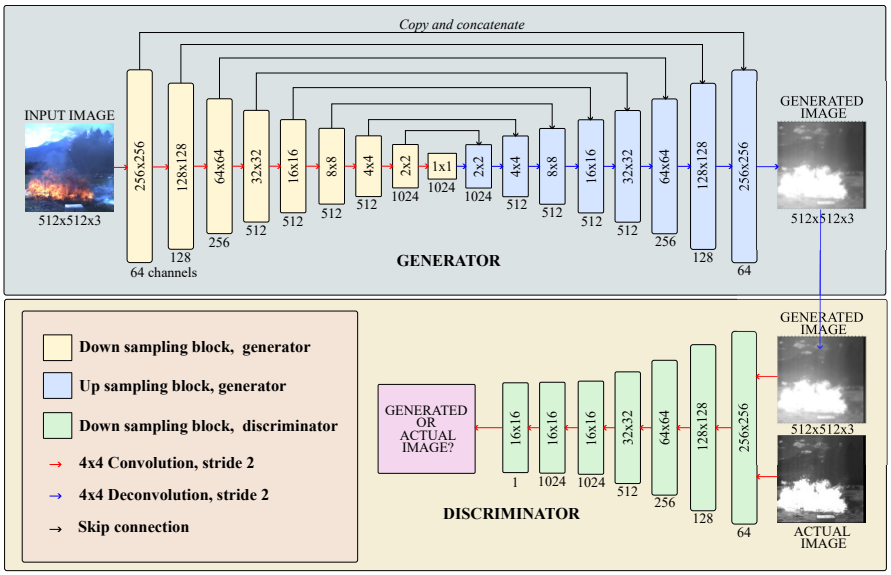

Figure 1: An example configuration of Pix2Pix as is used in our experiments.

tains pairs of aligned RGB and NIR images is required. The model is trained to minimize the difference between the generated NIR images and the ground truth NIR images in the dataset. This enables the model to learn the underlying patterns and correlations between RGB and NIR images, allowing it to generate plausible NIR representations when given an RGB input during the testing phase.

In our case, the tranformation from RGB to NIR image is approximated by a conditional generative adversarial network, which is introduced in the following section. This transformation is highly nonlinear which stems from many factors, such as lighting sources, surface reflections, intristic and extrinsic camera characteristics. Here, the challenge is to estimate the global optimal solution that guarantees convergence.

2.2 Pix2Pix Generative Adversarial Network

In this paper, we employ an existing image-to-image translation method known as Pix2Pix (Isola et al., 2017). Pix2Pix is a type of cGAN, which was introduced in 2017 by the researchers from Berkeley AI Research (BAIR) Laboratory, UC Berkeley. It can learn to map an input image from one domain to an output image in a different domain. Typically, Pix2Pix uses a paired dataset consisting of input images and their corresponding output images, while traditional GANs generate images from random noise. An example applications is the conversion of a black-and-white image into a colored image. Ideally, the network learns to generate realistic output images by training on these paired examples.

Similar to a traditional GANs, the architecture of Pix2Pix consists of two main components: a generator and a discriminator. In Fig 1. an example configuration of Pix2Pix is illustrated as is used in our experiments. In this paper, we investigate the use of Pix2Pix to understand the benefits the paired images can bring to this task. Below we briefly introduce the main components which are used for experiments.

2.2.1 Generator

The generator takes the input RGB image and attempts to transform it into an output NIR image. It is an encoder-decoder model with U-Net architecture (Ronneberger et al., 2015). This is accomplished by first downsampling (encoding) the input image down to a bottleneck layer, and then upsampling (decoding) the bottleneck representation to the size of the output image. In U-Net architecture, the skip-connections are added between the encoding layers and corresponding decoding layers, as denoted in Fig 1. The architecture consists of standardzed blocks of convolutional, batch normalization, dropout, and activation layers.

During training, the generator progressively improves its ability to generate realistic output images by minimizing the difference between the generated and real images, as perceived by the discriminator. It is trained via the discriminator model. The goal is to minimize the loss predicted by the discriminator for generated images which are predicted as real images.

2.2.2 Discriminator

The discriminator's role is to distinguish between the generated output image and the real output image from the dataset, thus the discriminator is trained on the real input and generated NIR images (Ronneberger et al., 2015; Schonfeld et al., 2020). The generator and discriminator are trained in an adversarial manner via adversarial loss, where the generator aims to produce output images that fool the discriminator, while the dis-

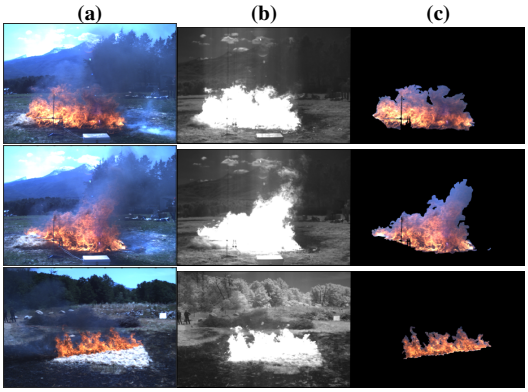

Figure 2: Example of images of the Corsican Fire Database taken with a multi-spectral camera. (a) visible (RGB) image, (b) Near Infrared (NIR) image, (c) ground truth flame pixels based on the visible image. Ground truth images are not used in this paper, however, they are shown here as a reference of what is a goal of detection from the original image.

criminator aims to correctly classify the real and generated images. The discriminator tries to become more accurate in distinguishing between real and generated images. This adversarial training process helps the generator learn to generate output images that are visually similar to the target domain. The update of the discriminator is unrelated to the update of the generator.

2.2.3 GAN

The generator and discriminator are connected together to create a composite model (Creswell et al., 2018; Goodfellow et al., 2020). Adam is used as an optimization method, with 0.0002 as a learning rate, while binary crossentropy and mean apsolute error (MAE) are used as loss funtions.

2.3 Dataset preparation

The Corsican Fire Database (CFDB) is used as our dataset (Toulouse et al., 2017). The CFDB comprises a collection of 500 visible images depicting wildfires captured worldwide, and 95 paired images consisting of visible and infrared modalities. These images are captured under realistic outdoor conditions. They are obtained using the JAI AD-090GE camera, which utilizes a 2-CCD multi-spectral camera system. The camera captures both visible and near-infrared spectra (700-900 nm) using the same optical setup. Both the visible and infrared images have dimensions of 1024×768 pixels. Each image in the dataset is accompanied by a corresponding ground-truth image. In this paper, the ground truth images are not used. Fig 2. illustrates some examples from the Corsican Fire Database. The CFDB dataset is openly accessible for research purposes.

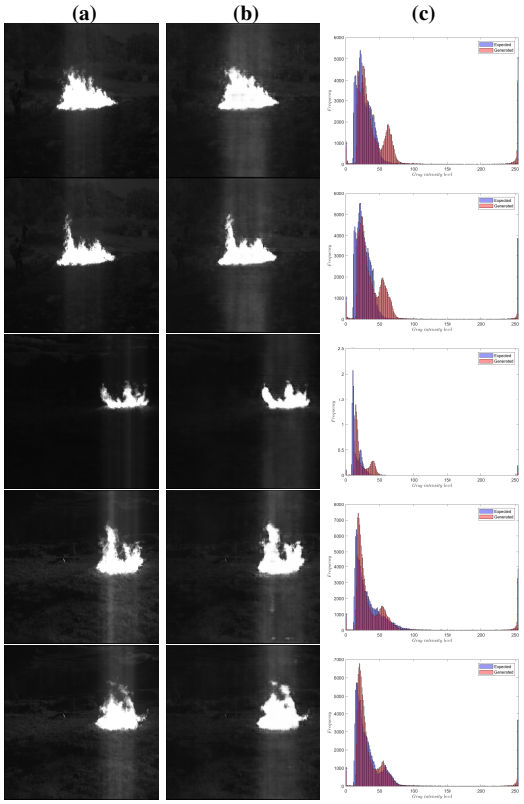

Figure 3: Results of our approach: (a) original NIR image from the CFDB, (b) generated NIR image, (c) and their histograms.

In this paper, we use 95 paired images from the CFDB, where 75 images are used (approximatelly 80%) during training, and the other images are used for testing purposes (20%). The images are first loaded and resized into a target size of 512×512 pixels to prepare them as inputs to Pix2Pix architecture.

2.4 The mapping procedure

The original RGB images are three-channels images, while NIR images are one-channel images. To learn the mapping function between RGB and NIR, we transform NIR images into three-channel images, where the content of the original channel is just copied to others. Then, we employ the classical Pix2Pix network where we experiment with diffrent configurations, and number of epochs and batch sizes. The generated images contain three channels, where information between channels differs slightly between 5-10%. To measure the performance of our approach, the generated NIR images have to be converted to the one-channel images. This procedure can be accomplished with several different methods, such as by computing the max, min, median or mean value of each pixel in three channels and then creating the one-channel images. We experiment with these results with respect to metrics and present the results in Section 3.

2.5 Metrics

The peak signal-to-noise ratio (PSNR) is used as a quality measurement between the original and a generated image (Korhonen and You, 2012). PSNR represents a measure of the peak error. The higher the PSNR, the better the quality of the generated image. To define the mathematical formulation, let us assume I and G are the original input and generated images, respectively. Then, the PSNR between I and G is given by Eq. (1).

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{I^2}{G} \right) \tag{1}$$

The Mean Squared Error (MSE) measures the average squared difference between the expected and the generated image of a dataset (Wang and Bovik, 2009). The lower the value of MSE, the lower the error. The MSE between I and G is given by Eq. (2).

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (I_i - G_i)^2 \tag{2}$$

The Root Mean Squared Error (RMSE) is calculated by taking the square root of the Mean Squared Error (MSE).

MSE and PSNR focus on measuring the absolute differences between the predicted and true values or the reconstructed and original images, without considering the perceptual characteristics of human vision. They provide a quantitative assessment of the fidelity or accuracy of the signals or images by considering the mean squared error or the signal-to-noise ratio.

On the other hand, SSIM (Structural Similarity Index) is a perception-based model that takes into account the structural information and important perceptual phenomena related to human vision (Wang et al., 2004). It considers the idea that pixels in an image are interdependent, particularly when they are spatially close. SSIM measures the similarity between two images by assessing the perceived change in structural information, taking into consideration luminance masking and contrast masking effects. The Structural Similarity Index (SSIM) is calculated as shown in Eq. (3).

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3}$$

3 Results and Discussion

The Pix2Pix model is trained from scratch using an available open-source implementation[1]. The training process consists of a fixed number of iterations, with 100 epochs and a batch size of 1. Given the presence of 75 images in the training dataset, each epoch comprises 75 iterations, resulting in a total of 7500 training steps for the entire process. During each training step, a batch of real examples is selected as the initial step. The generator is then employed to generate a corresponding batch of samples based on the input source images. Subsequently, the discriminator is updated using both the batch of real images and the batch of generated samples.

Afterwards, the generator model undergoes an update, wherein the real source images are used as input, and the expected outputs consist of the real target images, accompanied by class labels of 1 (indicating real images). The loss calculation is based on these inputs. The generator produces two loss scores, along with a weighted sum score. We focus on the weighted sum score, as it is instrumental in updating the model weights.

To monitor the training progress and ensure timely evaluation, images are generated every 10 epochs and subsequently compared to the corresponding expected target images. This approach provides a means of regularly assessing the advancement of the training process. Upon completion of training, the remaining 20 images in the test subset are utilized to generate NIR images, which are then compared with the original images. To gauge the quality of the generated images, four metrics are employed for the comparison between the original (expected) and generated images. The obtained results are presented in Table **??**.

Table 1: The metrics results for test NIR images from CFDB.

Metrics	Results
PSNR	21.95
MSE	516.09
RMSE	22.72
SSIM	0.8242

The mapping procedure from a three-channel to a one-channel NIR image involved determining the maximum, minimum, median, or mean value of each pixel across the three channels. After evaluating the results, it was found that the minimum function yielded the best outcomes. Hence, the minimum function was employed to generate the final metric results. It is worth noting that the disparity between the maximum, minimum, median, and mean functions was minimal, with variances falling within the range of 1-2%.

Fig. 3. displays a set of five original RGB images alongside their corresponding generated NIR images, accompanied by their respective histograms. The histograms provide clear evidence that the model has effectively learned the translation process from RGB to NIR, capturing the desired characteristics. Furthermore, the structural similarity index demonstrates a remarkable similarity between the generated NIR images

[1]Pix2Pix Github page: `https://github.com/phillipi/pix2pix`

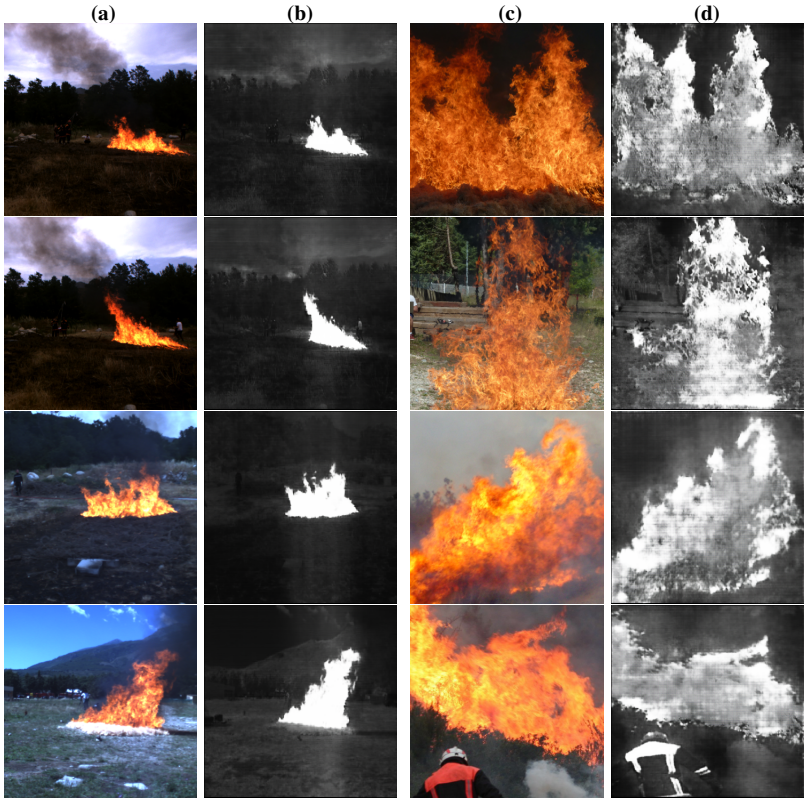

Figure 4: Results of our approach on a sample of 500 CFDB RGB images with satisfactory results: (a) original RGB image from the CFDB, and (b) generated NIR image; and with less satisfactory results: (c) original RGB image from the CFDB, (d) generated NIR image.

and the original ones, with an average value of 82.42%. Visual inspection confirms that the model excels in detecting the position and shape of the fire to a significant extent. However, it is worth noting that some details on the edges of the frames appear less prominent in the generated images.

The model was subsequently utilized to generate NIR images from the original 500 images from the CFDB dataset. In Fig 4., examples of both successful and less satisfactory translations are showcased. Since there are no corresponding original NIR images available for comparison, the evaluation is based solely on visual assessment. From the obtained results, it is evident that images resembling the ones encountered during training are effectively translated into their NIR counterparts, as seen in Fig 4. in (a) and (b). However, images that exhibit a higher density of fire pixels tend to have less satisfactory translations to their NIR representations, as seen in Fig. 4. (c) and (d). Although the model generally succeeds in accurately determining the position and shape of the fire, the internal structure appears blurred and requires additional processing.

To address this limitation, one potential approach would involve augmenting the training subset with similar samples and subsequently reiterating the training procedure. By incorporating a more diverse range of fire instances, the model can potentially improve its ability to capture finer details and enhance the translation of images with higher fire pixel density to NIR images.

4 Conclusion

In this paper, we demonstrate that the Pix2Pix model can be used to successfully perform image to image translation for the case of RGB to NIR translation. The model demonstrated its ability to successfully translate RGB images to NIR images, as evidenced by the clear similarities observed in the histograms and the high structural similarity index with an average value of 82.42%. The model effectively detected the position and shape of the fire, although some details on the edges of the frames were less pronounced in the generated images.

Applying the model to the CFDB dataset showed promising results for images resembling those encountered during training. The Pix2Pix model demonstrated its potential for generating NIR images and capturing key fire-related characteristics. Further refinements and augmentations could enhance its ability to produce high-quality NIR images in various scenarios.

References

Akagic, A. and Buza, E. (2022). Lw-fire: A lightweight wildfire image classification with a deep convolutional neural network. *Applied Sciences*, 12(5):2646.

Aslahishahri, M., Stanley, K. G., Duddu, H., Shirtliffe, S., Vail, S., Bett, K., Pozniak, C., and Stavness, I. (2021). From rgb to nir: Predicting of near infrared reflectance from visible spectrum aerial images of crops. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1312–1322.

Brdjanin, A., Dardagan, N., Dzigal, D., and Akagic, A. (2020). Single object trackers in opencv: A benchmark. In *2020 International Conference on INnovations in Intelligent SysTems and Applications (INISTA)*, pages 1–6. IEEE.

Brooksby, B. A., Dehghani, H., Pogue, B. W., and Paulsen, K. D. (2003). Near-infrared (nir) tomography breast image reconstruction with a priori structural information from mri: algorithm development for reconstructing heterogeneities. *IEEE Journal of selected topics in quantum electronics*, 9(2):199–209.

Buza, E. and Akagic, A. (2022). Unsupervised method for wildfire flame segmentation and detection. *IEEE Access*, 10:55213–55225.

Ciprián-Sánchez, J. F., Ochoa-Ruiz, G., Gonzalez-Mendoza, M., and Rossi, L. (2021). Fire-gan: A novel deep learning-based infrared-visible fusion method for wildfire imagery. *Neural Computing and Applications*, pages 1–13.

Clarke, F. (2004). Extracting process-related information from pharmaceutical dosage forms using near infrared microscopy. *Vibrational Spectroscopy*, 34(1):25–35.

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., and Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE signal processing magazine*, 35(1):53–65.

Dardagan, N., Brđanin, A., Džigal, D., and Akagic, A. (2021). Multiple object trackers in opencv: a benchmark. In *2021 IEEE 30th international symposium on industrial electronics (ISIE)*, pages 1–6. IEEE.

Dou, H., Chen, C., Hu, X., and Peng, S. (2019). Asymmetric cyclegan for unpaired nir-to-rgb face image translation. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1757–1761. IEEE.

Feng, C., Zhuo, S., Zhang, X., Shen, L., and Süsstrunk, S. (2013). Near-infrared guided color image dehazing. In *2013 IEEE international conference on image processing*, pages 2363–2367. IEEE.

Firmenichy, D., Brown, M., and Süsstrunk, S. (2011). Multispectral interest points for rgb-nir image registration. In *2011 18th IEEE international conference on image processing*, pages 181–184. IEEE.

Ghiass, R. S., Arandjelović, O., Bendada, A., and Maldague, X. (2014). Infrared face recognition: A comprehensive review of methodologies and databases. *Pattern Recognition*, 47(9):2807–2824.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, 63(11):139–144.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Kamruzzaman, M., ElMasry, G., Sun, D.-W., and Allen, P. (2012). Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis. *Analytica Chimica Acta*, 714:57–67.

Kapo, M., Šabanović, A., Akagic, A., and Buza, E. (2023). Early stage flame segmentation with deep learning and intel's openvino toolkit. In *2023 XXIX International Conference on Information, Communication and Automation Technologies (ICAT)*, pages 1–6. IEEE.

Korhonen, J. and You, J. (2012). Peak signal-to-noise ratio revisited: Is simple beautiful? In *2012 Fourth International Workshop on Quality of Multimedia Experience*, pages 37–38. IEEE.

Liu, M.-Y., Breuel, T., and Kautz, J. (2017). Unsupervised image-to-image translation networks. *Advances in neural information processing systems*, 30.

Lv, S., Huang, X., Cheng, F., and Shi, J. (2022). Color restoration of rgb-nir images in low-light environment using cyclegan. In *3D Imaging - Multidimensional Signal Processing and Deep Learning: 3D Images, Graphics and Information Technologies, Volume 1*, pages 1–9. Springer.

Matsui, S., Okabe, T., Shimano, M., and Sato, Y. (2011). Image enhancement of low-light scenes with near-infrared flash images. *Information and Media Technologies*, 6(1):202–210.

Mishra, P., Passos, D., Marini, F., Xu, J., Amigo, J. M., Gowen, A. A., Jansen, J. J., Biancolillo, A., Roger, J. M., Rutledge, D. N., et al. (2022). Deep learning for near-infrared spectral data modelling: Hypes and benefits. *TrAC Trends in Analytical Chemistry*, page 116804.

Pang, Y., Lin, J., Qin, T., and Chen, Z. (2021). Image-to-image translation: Methods and applications. *IEEE Transactions on Multimedia*, 24:3859–3881.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer.

Rüfenacht, D., Fredembach, C., and Süsstrunk, S. (2013). Automatic and accurate shadow detection using near-infrared information. *IEEE transactions on pattern analysis and machine intelligence*, 36(8):1672–1678.

Sa, I., Lim, J. Y., Ahn, H. S., and MacDonald, B. (2022). deepnir: Datasets for generating synthetic nir images and improved fruit detection system using deep learning techniques. *Sensors*, 22(13):4721.

Šabanović, A., Ahmetspahić, N., Kapo, M., Buza, E., and Akagic, A. (2023). Early stage flame segmentation with deeplabv3+ and weighted cross-entropy. In *2023 XXIX International Conference on Information, Communication and Automation Technologies (ICAT)*, pages 1–6. IEEE.

Salamati, N., Germain, A., and Siisstrunk, S. (2011). Removing shadows from images using color and near-infrared. In *2011 18th IEEE International Conference on Image Processing*, pages 1713–1716. IEEE.

Schaul, L., Fredembach, C., and Süsstrunk, S. (2009). Color image dehazing using the near-infrared. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 1629–1632. IEEE.

Schonfeld, E., Schiele, B., and Khoreva, A. (2020). A u-net based discriminator for generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8207–8216.

Shukla, A., Upadhyay, A., Sharma, M., Chinnusamy, V., and Kumar, S. (2022). High-resolution nir prediction from rgb images: Application to plant phenotyping. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 4058–4062. IEEE.

Sun, T., Jung, C., Fu, Q., and Han, Q. (2019). Nir to rgb domain translation using asymmetric cycle generative adversarial networks. *IEEE Access*, 7:112459–112469.

Toulouse, T., Rossi, L., Campana, A., Celik, T., and Akhloufi, M. A. (2017). Computer vision for wildfire research: An evolving image dataset for processing and analysis. *Fire Safety Journal*, 92:188–194.

Wang, Z. and Bovik, A. C. (2009). Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, 26(1):98–117.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612.

Yan, L., Wang, X., Zhao, M., Liu, S., and Chen, J. (2020). A multi-model fusion framework for nir-to-rgb translation. In *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pages 459–462. IEEE.

Yuan, X., Tian, J., and Reinartz, P. (2020). Generating artificial near infrared spectral band from rgb image using conditional generative adversarial network. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:279–285.

Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., and Shechtman, E. (2017). Toward multimodal image-to-image translation. *Advances in neural information processing systems*, 30.