EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Colorizing Multi-Modal Medical Data: An Autoencoder-based Approach for Enhanced Anatomical Information in X-ray Images

Bunny Saini [1], Divya Venkatesh[2], Avinaash Ganesh[3], Amar Parameswaran[4], Shruti Patil[5,6,*], Pooja Kamat[7,*], Tanupriya Choudhury[8]

[1,2,3,4,5,7] Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, 412115, Maharashtra, India
[6] Symbiosis Centre for Applied Artificial Intelligence (SCAAI), Symbiosis International University, Near Lupin Research Park, Gram: Lavale, Tal: Mulshi, Pune, 412115, Maharashtra, India
[8] CSE Dept., Graphic Era Deemed to be University, Dehradun, 248002, Uttarakhand, India

Abstract

Colourisation is the process of synthesising colours in black and white images without altering the image's structural content and semantics. The authors explore the concept of colourisation, aiming to colourise the multi-modal medical data through X-rays. Colourized X-ray images have a better potential to portray anatomical information than their conventional monochromatic counterparts. These images contain precious anatomical information that, when colourised, will become very valuable and potentially display more information for clinical diagnosis. This will help improve understanding of these X-rays and significantly contribute to the arena of medical image analysis. The authors have implemented three models, a basic auto-encoder architecture, and two combined learnings of the autoencoder module with transfer learning of pre-trained neural networks. The unique feature of this proposed framework is that it can colourise any medical modality in the medical imaging domain. The framework's performance is evaluated on a chest x-ray image dataset, and it has produced benchmark results enabling high-quality colourisation. The biggest challenge is the need for a correct solution for the mapping between intensity and colour. This makes human interaction and external information from medical professionals crucial for interpreting the results.

Keywords: Gray Image Colorization, Unsupervised Learning, Auto Encoders, Transfer Learning, Medical Imaging

Received on 20 December 2023, accepted on 17 March 2024, published on 25 March 2024

doi: 10.4108/eetpht.10.5540

*Corresponding author. Email: pooja.kamat@sitpune.edu.in

1. Introduction

The human eye perceives colour changes better than shades of grey, so one must colourise images to make their interpretation and understanding easier. Colour images contain more detailed and clearer information compared to grayscale ones. In many areas, only grayscale images are generated. Therefore, this grayscale image can be transformed into a colour image to better visualise it. Colourisation is converting a grayscale image or video into a coloured one. This is executed by interpreting its luminance or intensity values and assigning them values of red, green, and blue colour channels to generate various colours. During colourisation, a vector in the specified colour space used to represent a particular colour is added in place of the old scalar value that represents the pixel's intensity [1]. Some of the most popular colourisation techniques are based on: luminance keying and colour transfer, image analogies, motion estimation, segmentation, colour prediction, probabilistic relaxation, and chrominance blending [2].

Medical imaging is imaging of various parts of the human body, such as bones, organs, and tissues that are later used for medical investigation and clinical treatment for patients. Many different imaging methods are used to examine diseases like X-rays, CT scans, MRIs, and Ultrasounds [3]. These images help pinpoint the problem areas to provide better patient treatment. Modern imaging techniques enable examining complex internal biological processes otherwise impossible to visualise perfectly, like mutation, metabolism, blood circulation and chemical reactions. With the recent advancements in science, medical imaging is contributing in huge ways to disease diagnoses and understanding the human anatomy [4]. But the noteworthy point here is that almost all medical image forms are inherently grayscale. The authors are focussing primarily on the X-ray images in this study. The process followed during X-ray imaging utilises fundamental physical phenomena. It captures the stretching caused by the acoustic wave dissemination of X-ray propagation through the body parts to perceive and understand the body part's health. Another versatile contribution of X-rays is robotic-assisted surgery's guided surgical assistance [5].

Colourising X-rays is essential as adding colours to images increases the visibility of even a small bone or tissue. The main objective of this paper is to colourise medical images to enhance all the important information in them. The colourised medical image mustn‘t distort any features present in the input image. This should be performed carefully, as any extra synthesised feature added to the image can cause misdiagnosis [6]. Good visual quality and the complete chromatic information in medical images significantly improve analysis and diagnosis.

This paper aims to develop a fully automatic colourisation mechanism that inputs a grayscale image and outputs a colourised image. Much such work has been successfully implemented using deep neural networks (DNNs). Most previous work on this topic required training examples for both coloured and grayscale image versions. However, obtaining such data is expensive, difficult, and sometimes even impossible for the current problem statement. In this paper, the proposed models use only grayscale image data for training. The authors attempt to compensate for the lack of coloured images by applying transfer learning[7] knowledge from domains where the information collected from coloured data is abundant. The goal is to use existing deep neural network architectures like ResNet [8]. These architectures have been pre-trained on one of the most famous and diverse image datasets, ImageNet [9]. The authors plan on simply fine-tuning their parameters and adapting them for use with the dataset of medical images. This gives us the benefit of learning from scratch while also using the learned information of pre-trained source models by focusing on colour adaptation [10]. This approach is very efficient as the data available is scarce. The method enables us to transfer the newly learned colour information across multiple medical datasets.

This paper is split into sections, each of which lays the groundwork for the next and establishes the paper's structure. Section I Introduces us to the problem this study aims to address in the next section, Section II: A Literature Review. Section III details the data acquisition phase, followed by Section IV, which gives us a quick rundown of the work methodology and process flow. Section V provides elementary knowledge of all the models utilised in the process. Section VI: Evaluation Metrics used to assess, and Section VII highlights the results and discusses the key takeaways from this study. The paper concludes with Section VIII: Conclusion and Future Scope.

2. Related Work

According to the research trends in the past few years, Deep Learning Image Classification has become more dependent on colour space. Colour space, which characterises colour from individuals' subjective sensations to tangible expression, is the theoretical basis of colour information studies. It also offers a strong foundation for image capturing and colour performance. The notable RGB colour space is often employed in image processing. Red, Green, and Blue are the channels that make up this colour space [11]. Each channel contains 256 grey levels, and multiple colours may be produced by combining the channels in a certain way. The most used colour space for line art picture colourisation is RGB. Some of the other spaces are CIELAB (any colour can be expressed independently) [12] and YUV (depending on the brightness and chrominance) [13]. Different colour spaces are compatible with various colourisation techniques. Most researchers choose a solution that is extremely practical and consistent with their methodology, while others examine how different colour schemes interact with their tasks.

The broad division of models that have been implemented for the task are (i) Deep Convolutional Networks, (ii) Generative Adversarial Networks (GANs) based architectures, and (iii) Transformer-based methods.

Colourisation methods utilising Deep Learning divide the reference photos into several groups using the suggested adaptive graph clustering technique and deploy a colourisation network separately for each image cluster stack. Notably, one of the fundamental efforts in the domain of Zhang, R (et al.) developed a deep CNN and a carefully designed objective function that might help colourisation get closer to providing outcomes identical to actual colour photographs. After deleting colours from the input, classifier performance fell from 68.3% to 52.7%, but their whole system fooled participants in 32% of trials. Utilising their whole process, decolourization boosted the performance to 56%.[14] Cevallos S. (et al.) proposed a customised deep CNN model and a modification of the VGG16 model, resulting in a deep convolutional auto-encoder framework that yields promising results with little quality loss, which has also been adapted in this study [15]. Dong X. (et al.) suggest that direct training from camera systems is possible, and it doubles the colourisation using Weighted Average Colorization. They propose a Global Curve Adjustment network that reduces losses and guarantees spatial smoothness in the second colourisation output, producing superior results [16]. Kim, S (et al.) proposed a volume rendering method based on images that mirror the desired image. TF grid labels were determined using a CNN model. They extract the main colours from the target image using the CNN model. For this framework, VGGNet performs computations more quickly than ResNet and AlexNet. However, a larger experiment is needed since the experiment was only evaluated using a few datasets. [17]. Shi, Q (et al.) describe a VGG-16 Convolutional network model that uses a fully autonomous learning-based colorisation technique based on cross-entropy loss in classification generating plausible colour images while mainly focusing on reducing image quality loss [18]. Joshi, M. R. (et al.) looks into the automated colorisation of historical, cultural, and legacy pictures now only available in monochrome. It uses the Inception-ResnetV2 architecture with back-propagation support to identify patterns in RGB and grayscale information. The neural network is deployed to identify a* and b* channel colour values provided by the L grayscale input. Overfitting was not discovered since the model incorporates the semantic data from the inception module. Additionally, several data augmentation strategies reduced overfitting. From the standpoint of increased signal energy, their framework provides image colourisation with better or equivalent results compared to several existing techniques [19]. Mouzon, T (et al.) proposed a new approach to combine CNN's strength with the precision of the variational models. This integration is accomplished by broadcasting data that depends on probability distributions. The computation of the two framework components is based on accepted methods published in the research domain. The numerical outcomes demonstrate how colourisation outcomes improved when the two procedures were combined [20]. Nguyen T. (et al.) provided a technique for colouring pictures that combines the colour and style of a colour image with the content of a grayscale image to colourise it. This technique employs a CNN to extract colour information from one image and transfer it to another. They displayed instances of convincing-appearing generated images. They deploy L-BFGS, a quasi-Newton approach that closely resembles the BFGS algorithm, to increase optimisation efficiency and obtain better-looking pictures [21].

The key distinctions in GANs-based colorisation models are often found in the network architecture, specifically the loss function and the generator structure. To make these models more suitable for colouring images, the bulk of these approaches introducing colour depends on conventional generative models. Hensman, P. (et al.) postulates that high-level structures like hair, skin, and clothing are identified and given the appropriate colour in the cGAN colorization stage in a translation-invariant way. Their post-processing techniques reduce any remaining errors. The Results are improved by training the cGAN on a cropped copy of the training picture that corresponds to the target image. A segmentation and colour-correction approach reduces blur, ambiguities, and other issues in the output. The results of the cGAN colorization are frequently good even when used alone and trained quickly [22]. Mourchid, Y. (et al.) proposes GANs based architecture that employs two discriminators for features in addition to content picture and style transfer methods that transfer the textures and details of an image. The experimental results show that the proposed model can dependably produce stunning colourized images with fewer artefacts than state-of-the-art methods [23]. Li, Z. (et al.) suggest a novel HistoryNet design that includes classifier, parser, and classification subnetworks inspired by InfoGAN [24]. A semantic parsing subnetwork can aid in improving the accuracy of the colorization border. A classifier subnetwork can aid in selecting the proper colour. Their technique outperforms the cutting-edge colorization network in LPIPS, PSNR, and SSIM through both qualitative and quantitative comparison [25].

For transformer-based methods, a conditional self-regression is used as the initial step in the procedure. The obtained image is first transformed into a low-resolution, coarsely coloured image, and furthermore, it is upsampled using two networks that are entirely parallel to produce the finely coloured images of high-resolution. To address the colorization of historic black-and white aerial photographs, Farella E. M. (et al.) introduce a novel neural network architecture called Hyper-U-NET, which combines a U-NET-like design and HyperConnections. Although there were occasional failures when working with low-quality images, the suggested approach generally produced excellent colorization results in many cases [26]. Dias, M. (et al.) assessed the effectiveness of a modified W-Net design. Their W-Net model has been taught to separate pictures while reconstructing or predicting the colour of the input images from intermediate representations. They demonstrate how well the suggested W-Net architecture colours and segments the input pictures. The suggested method surpasses some of the previously presented methods when segmenting both coloured and grayscale images, as well as a baseline according to the U-Net model. [27].

Several survey literature was referred for this study to analyse the applications and implementation efforts in the domain. Chen, S. Y. (et al.) describes how most systems for colouring sketch images combine completely automatic colorization models with human assistance since little adjustments can improve the outcome and bring it closer to the user's expectations. Since most anime and manga character colours are unknown, using the "wrong" colour won't be a major setback while colouring sketches. The interactive colorization techniques based on cues, sketches, or text can only give a region's colour prior knowledge; they cannot alter the region's boundaries [28]. Jin X., (et al.) outlines that Deep Learning Image Colorization models are categorised due to their description of the DLIC issue specification and their introduction of the most often used colour space and loss function. Then, novel concepts are put forth that consider the network topology, level of automation, and application domain. These techniques can successfully colour some specialised grayscale images, but the bulk of those now in use still need help with colour overflow, uneven colouring, and unsaturated tones [29]. Žeger, I. (et al.) summarises how because of discrete components of its neural network architecture that were trained for the local characteristics (colour distribution) and global data, Zhang et al. 's automated version produces better results (semantic information). They perform an extensive survey of User Based Scribble Methods and Deep Learning Methods. Due to a stronger ability to adapt to the colorization challenge, more complicated structures, like GANs in the diverse category, produce believable outcomes. The user-guided colorization neural network outperforms the scribble-based technique in terms of colorization among methods needing user input [30].

For the past decade, colorization of medical pictures has been an active area of research. The use of coloured medical pictures facilitates disease prevention and treatment. Colourizing medical images is suggested to aid medical professionals in the automated detection of affected areas and a better understanding of the region of interest. Nida, N. (et al.) proposes a system that computes chromatic values of the target reference medical photographs based on their brightness, and then applies them to the needed source image to offer colorization. Performance of the proposed system yields benchmark results. It was discovered that the outcomes were essentially the same as the original input image, but with far more chromatic and visual information [31]. Lagodzinski, P (et al.) discuss an approach for morphological distance transformation that considers the Computed Tomography (CT) Scan picture structures and perpetuates user-attributed colour to the image automatically. This method yields good results for many grayscale medical images while retaining the original image's intensity. A healthcare professional may be aided by the colour information as a tool to highlight significant structures by colouring the region of interest [2]. Mathur, A. N. (et al.) offers a framework with the intention that it might be applied to multi-modal data in 3D for colorisation while keeping in mind the photorealism of the resulting outputs. They demonstrated the usage of a synthetic dataset to build an alternative modality that is much more suited to visual perception, then utilising the generalisation abilities of a GAN to expand this strategy on unseen data. Then, designed a system that could support colorization from contextual photos and be expanded to 3D. Additionally, they illustrate the multi-modal and three-dimensional nature of medical imaging data causes some state-of-the-art 2D algorithms to significantly fail. All these factors result in robustness in terms of metrics and visual perception [32]. Gotoh, Y. (et al.) advocates the use of preprocessing steps taken to improve edge information and decrease noise. Each pixel of a subject image is then given colour values. Pulse coupled neural networks along with thresholding are utilised to segregate differentiated images. The resultant picture may represent anatomical information more precisely than a grey image. Their approach works better than many other methods because of the image refining steps prior colorization [4]. Liang, Y. (et al.) extended CycleGAN introducing perceptual as well as TV loss to it. Their strategy adapts a considerable advantage of training the model without requiring paired instances. Incase of no matching training data, this feature is quite useful, especially for medical images.Despite having a similar training and testing complexity as CycleGAN, their model outperforms it and yields better outputs [6]. Selvapriya, B. (et al.) suggests that the binary image of grayscale input is defined using the Region of Interest for which separate RGB channels are extracted and designated. Output image is displayed after merging individual channels into a single image. The primary benefit of their work is that, rather than colouring the entire image, they have chosen the specific area that needs medical attention and introduced colours for that Region of Interest [33]. Lamberti, F. (et al.) proposes in their technique, a deep colorization module is used to discover the best colour mapping for X-ray Images using a given pre-trained CNN. They suggest a multi-stage transfer learning process in which the lightweight colorization module is first taught from scratch while the backbone is kept frozen, and both modules are then fine-tuned toward the target goal in a refinement step. Their method results in a sizable performance improvement when the feature encoder is frozen over two datasets and two distinct backbones [10].

3. Dataset Information

Chest X-Ray images is a publicly accessible dataset obtained from Kaggle [34]. It includes grayscale X-Ray pictures of both normal and pneumonic patients. The training, testing, and validation portions of the dataset are organised into three folders. There are 5863 X-Ray pictures in all, divided into two groups (Pneumonia/Normal).

The chosen chest x-ray images were taken at the Guangzhou Women and Children's Medical Center of paediatric patients in the age range of one to five. These patients received regular clinical treatment, which included all chest X-ray imaging. All chest radiographs were originally inspected for quality control prior to the processing of the X-ray pictures, and any scans that were of poor quality or that couldn't be read were removed. The classification of images containing Pneumonia or not was done by two expert physicians before being cleared for training the AI system. A third expert also overlooked these classification and re-evaluated them later in order to account for any misclassification.

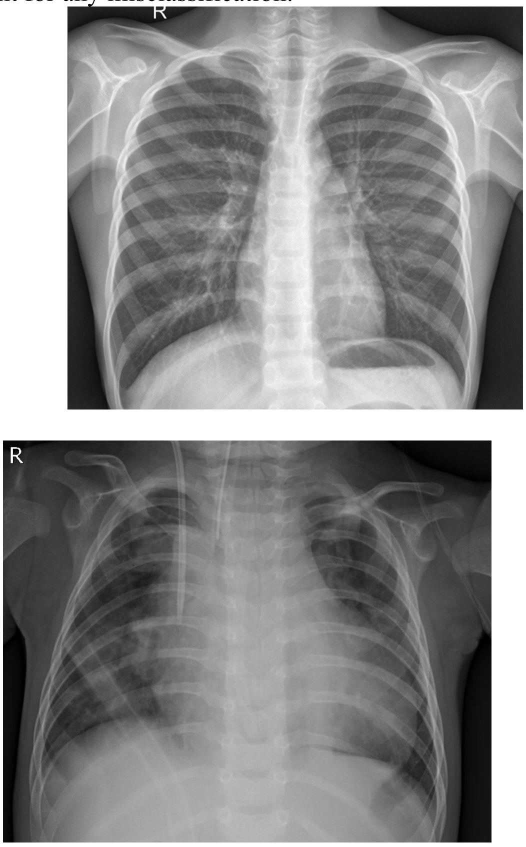


Fig 1. Normal vs. Pneumonitis Chest X-Ray Images from [34]

4. Methodology

Most visual tasks can be effectively performed using Convolutional Neural Networks (CNNs). Using deep learning models on visual tasks has enabled us to see rapid strides in their produced outputs and applications. Modern deep learning models are trained on large image datasets within a reasonable time and hardware provides consistently good results [29]. A relatively large dataset is required during training as fewer images will not provide sufficient colour information to be able to produce satisfactory colorization results. In this case, the desired outcome would be for the model to be able to learn colours automatically and yield results that correspond reasonably to real-life objects and their colour. For this procedure, the authors use network architectures called autoencoders and train them using transfer learning methodology. Subsequently, the authors explain the implementation procedure and the steps undertaken to complete this work.

The main steps of the procedure followed, as shown in Figure 2 are:

1. A dataset which had chest x-rays of patients with and without pneumonia was chosen to be colourized in the research work. This dataset was then downsampled to lower the computation power required in the training process. The final dataset that was used for training the models contained 400 train images (200 each of normal and pneumonia) and 100 test images (50 each of normal and pneumonia).
2. This step was followed by data preprocessing procedures. Here, the authors resize the images to correctly align them with the chosen autoencoder architecture. It was adjusted such that the image size perfectly matched the input dimension of the first layer of the network.
3. The next step was to load the data into the chosen model for training. Two of the three models also execute transfer learning using ResNet-13 and ResNet-34 on that model. The architecture is divided into two parts: an encoding layer and a decoding layer. Out of this, the encoding layer is the one that is trained using transfer learning.
4. Next, the model makes the colour predictions for the x-ray images it was trained on using the learned weights and biases. Here, a colourised version of each x-ray gets stored on the device cloud and can be used to visualise the results achieved by the model.
5. Conclusively, this colourised x-ray image output can be used to determine if a patient is suffering from pneumonitis. Once the model is created and adjusted to fit the requirements, it is compiled and stored to be simple to retrieve in the future.

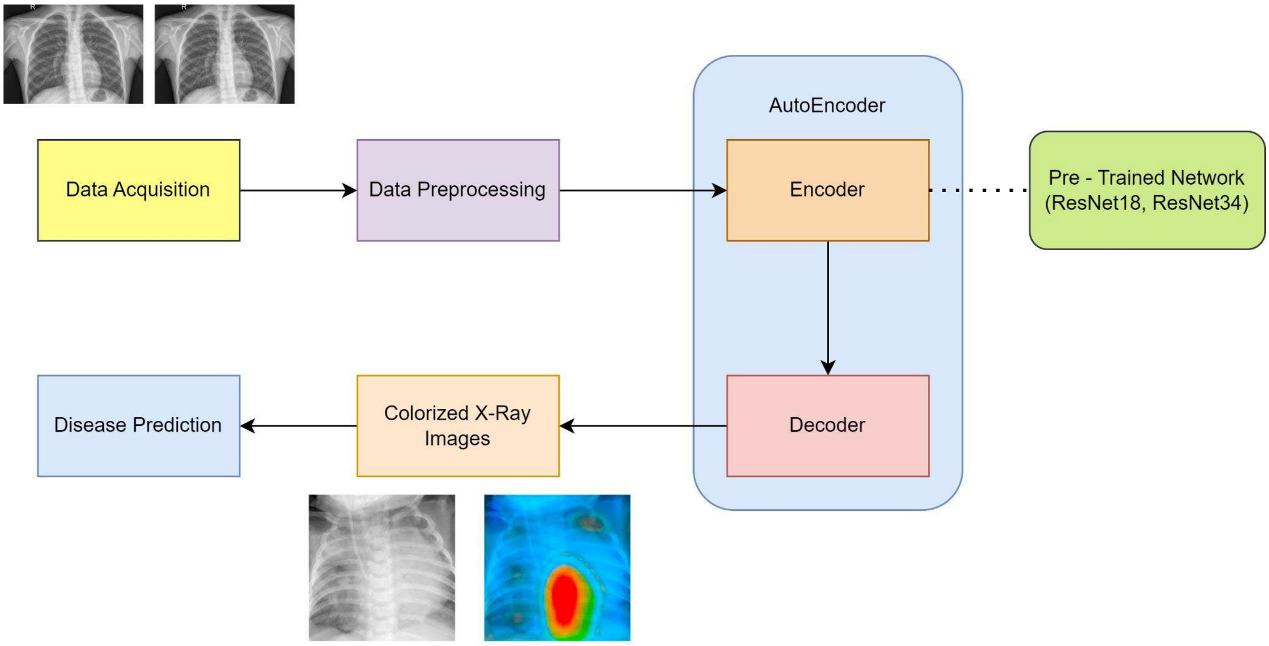


Fig. 2 Process flow diagram for this study

5. Overview of Deep Learning Models

This section is divided into three parts for the three colourization methods that the authors have implemented - Autoencoders and two U-Net deep learning models (ResNet18 and ResNet34).

5.1 Autoencoders

Autoencoders are a type of artificial neural network that via back propagation learn to reconstruct data by effectively compressing and encoding the data into meaningful representations first. They then proceed to reconstruct the data. This reconstructed data is meant to be as close to the original input as possible with the exact same dimensions. Autoencoders are used extensively in image processing and colorization applications. Their main objective is to learn the useful representations in the image with little or no supervision [36]. They consist of two neural networks, one for each encoder and decoder respectively, that are stacked together to form a deep network architecture of convolutional autoencoders. This network has proven particularly effective for anomaly detection where the objective is to identify the data samples that are not conforming to the original set. These anomalies are learned from a normal profile while being given only the normal data examples [37]. This is achieved by learning the latent subspace of normal samples. The goal is to produce similar results for colorization using the same properties of the network as mentioned above.

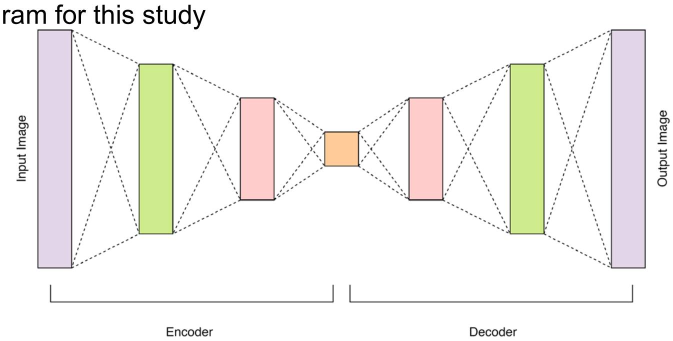


Fig. 3 Architecture of an AutoEncoder

An Autoencoder consists of three major components:

- Encoder: The input is compressed into a latent space representation in this section of the network. The encoder layer compresses and reduces the dimensions of the input image for encoding. After it is entered into the encoder, the original image gets converted into its compressed form [38].
- Bottleneck: The bottleneck is a key attribute of the network architecture. The bottleneck prevents the network from only memorising the input values without identifying its essential features. It places restraints on the amount of information that can traverse through the network, forcing a learned compression of the input data. The bottleneck is used to learn the attributes of the image from the compressed data during training [39]. This is the hidden layer present after the encoder structure which is used to optimise and only learn the significant features effectively.
- Decoder: This layer restores the image's original dimensions once it has been encoded. The original picture is lossy-reconstructed in the decoded image using the latent space representation. It decompresses the knowledge

representations learned during the bottleneck phase. This knowledge is then used to reconst the data back from its encoded form. The final output obtained from this layer is then compared with the ground truth [40].

5.2 U-Net

U-Net is a convolutional neural network architecture that was initially proposed for image segmentation. The U-Net network works by adding a layer-by-layer connection between each of the encoder and decoder layers. This forms somewhat of a U-shaped connection structure between them, resulting in its name [30]. In this work, the authors are using a residual U-Net (RU-Net) based autoencoder model that has the pre-trained ResNet architecture in place of the encoder network. ResNet networks are commonly applied as the backbone for input image feature extraction. The concept of skip connection between the deep and shallow networks of residual architecture is introduced into the network structure. This helps in avoiding saturation and even an unequal drop in accuracy across the network layers. Additionally, the ResNet series of networks coordinate the various feature graph sizes between each layer and serve as an efficient input for later networks [37]. Results obtained from ResNet networks of different depth, as expected, vary. ResNet depths adopted in this paper are 18 and 34, both yielding varied results which can be seen in later sections. More about the ResNet networks used in this paper are discussed below.

5.3 ResNet18

The ResNet18 is a pre-trained, 18-layer deep convolutional neural network model for PyTorch. It has been trained on the extensive and exhaustive images contained in the ImageNet database. There are eighteen layers in the network out of which 17 are convolutional layers and it is completed using a final fully-connected layer. The fully-connected layer followed by an average pooling layer at the end. The network includes residual shortcut connections that are inserted throughout between layers [41].

5.4 ResNet34

The ResNet34 is a pre-trained, 34-layer deep convolutional neural network model for PyTorch. It has been trained similarly on the images from the ImageNet database. Training deep neural networks can be challenging. ResNet 34 is a deeper network than the previous version but is easy to train compared to other commonly used deep neural networks due to its residual training framework. The considerably increased depth also gives us higher accuracy and solves the degradation problem. The 34-layer ResNet is better than the previously used 18-layer ResNet as it considerably lowers training error. It was also found to be generalisable to the validation data without any additional changes [42].

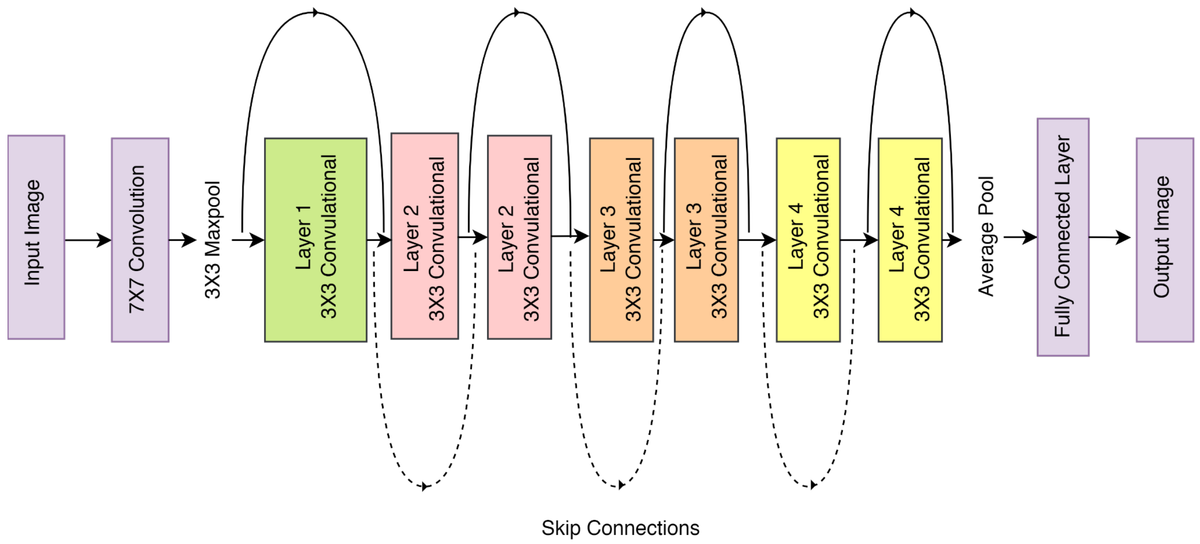


Fig. 4. Architecture of ResNet-18

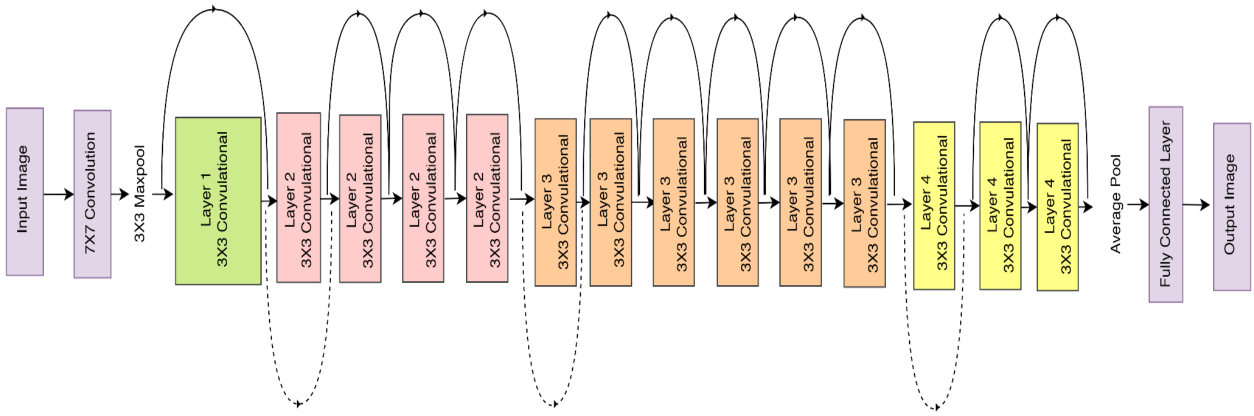


Fig. 5 Architecture of Resnet-34

6. Evaluation Metrics

This study's primary goal is to show the coloured outcomes of a grayscale image colorization process, with quality metrics serving as a supplementary scale of assessment. Choosing a decent evaluation criteria is challenging since colouring is very subjective and there are many acceptable variations. Throughout the course of this study, the authors discovered that the nature of the task implies some unusual observations: since colorization is multimodal problem - a single grayscale image may correlate to multiple admissible colour images, and since medical images are often acquired in monochrome, there may be no existence of ground truth. Because of this, human inspection or in the given context, medical assessment is frequently necessary. Using the Mean Squared Error (MSE) as a loss function is a fundamental method for training Neural Networks for the task of colorization as utilised in the research work ([14], [43], [44], [45], [46]).

The advantage of employing this loss function is that it is simple to build, allowing us to concentrate on network architectures with more elaborate patterns. The authors can reduce the gap between the anticipated outcome and the intended output by using MSE loss. One of this loss' primary limitations is that it encourages conservative predictions, which has little influence on the work at hand since there is no ground truth for the colourized versions, because the authors concentrate more on the area the model has designated as a region of interest. Since L2/MSE determines the chroma in Euclidean space and effectively reduces MSE in this space, it proves to be an appropriate loss function for CIELab colorspace [44]. The equation for Mean Squared Error (MSE) is given by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{Y_i})^2 \qquad (1)$$

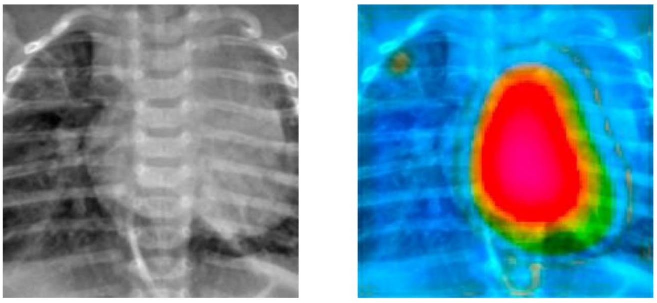
n = number of data samples
Y_i = observed values
$\hat{Y_i}$ = predicted values

7. Results and Analysis

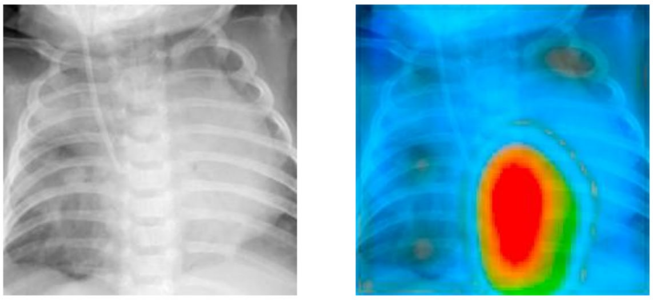
The computational setup consisted of a local system PC and cloud delivered Google Colaboratory PC. The local PC runs on an Intel i5 9300H @ 2.8 GHz (up to 4.0 GHz) with 4 Cores, 8 Threads and 8 GB DDR4 Memory. Along with a GTX 1650Ti GPU with 1024 CUDA cores and 4 GB Memory. On Google Colaboratory the authors were given access to a PC running on the Xeon Processor @ 2.3Ghz having 1 Core and 2 Threads with a Tesla K80 GPU, Compute 3.7 with 2496 CUDA cores and 12 GB Memory. Each iteration takes about 1.8-2 seconds of processing time.

The proposed method offers noticeably better visual understanding of medical pictures. The model favours paying close attention to and carefully evaluating the affected areas of X-rays revealing evidence of pneumonitis. Images that lack symptoms have a layer of neutral tone across the whole region, not emphasising any one portion of the grayscale representation. The output of the framework shows that anatomical characteristics become more apparent once the complete network is processed, illustrating pre pre-trained encoder network is slightly unable to detect essential visual aspects and that extra information on shape and texture is vital along with overall colour [47]. Without any human assistance or intervention, the pretrained networks effectively added a colour channel to the concerned patches and produced correct colourized versions of the monochrome images. However, a fundamental AutoEncoders structure was unable to recognise the data, failing to deliver results that were satisfactory.

The authors emphasise that the goal of this study is to enable the network to select the colours that will maximise identification accuracy rather than to produce images that are legible by humans. The future scope of this work could involve examining under which circumstances and how much colorization highlights the regions that are crucial for the final diagnosis which could be enabled with the support of a healthcare professional.

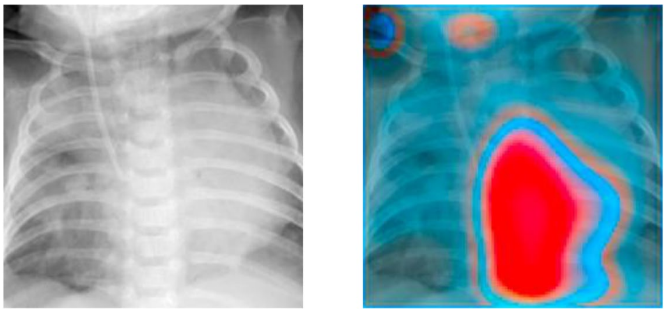


(a) Monochrome images

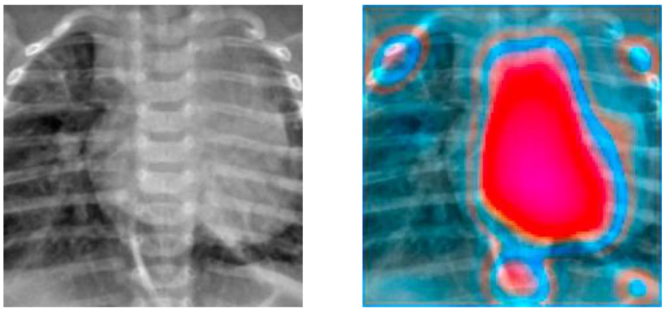


(b) Colourised Images

Fig. 6 Visual Representation of Monochrome vs. Colourised Images by network utilising Resnet-34



(a) Monochrome images



(b) Colourised Images

Fig. 7 Visual Representation of Monochrome vs. Colourised Images by network utilising Resnet-18

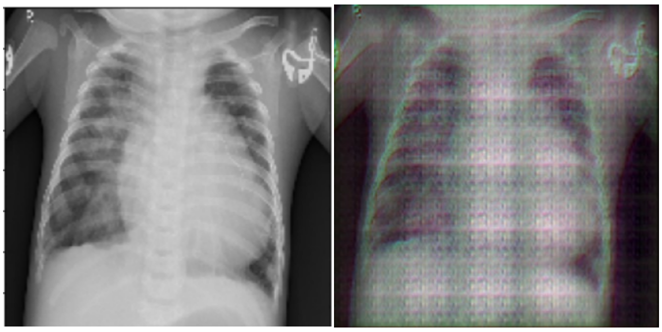


Fig. 8 Visual Representation of Monochrome vs. Colourised Images by Standard AutoEncoder

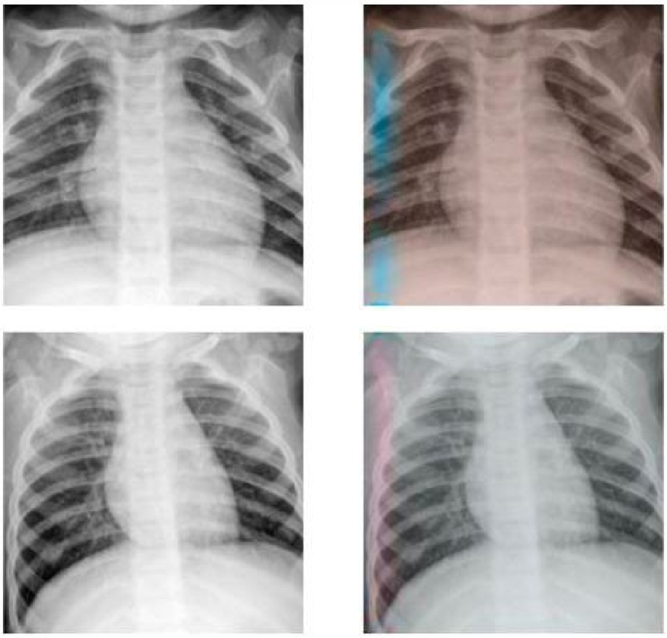


Fig. 9 Images with no signs of Pneumonitis depicting a plain tone overall, no colour information

In general, the loss graphs for this study display a declining trend with a few unstable peaks. It is observed that the majority of the models exhibit the best loss minimization during the second or third iteration itself and then go on to exhibit some overfitting in the following epochs. While the training loss shows some disruption, presumably due to overfitting on the X-ray data, the loss for the validation set of the pictures reduces to a minuscule amount in a sharp downward slope.

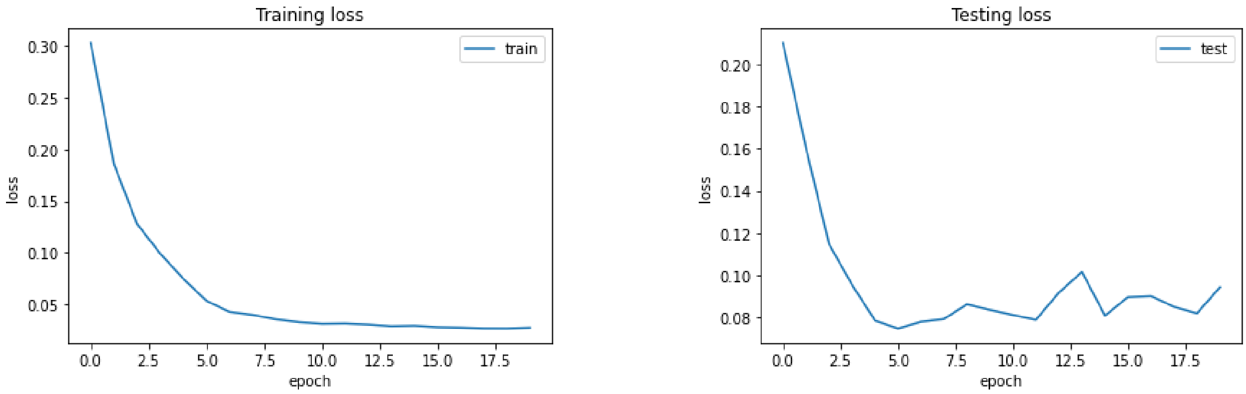


Fig. 10 Training vs. Validation Loss for Standard AutoEncoders

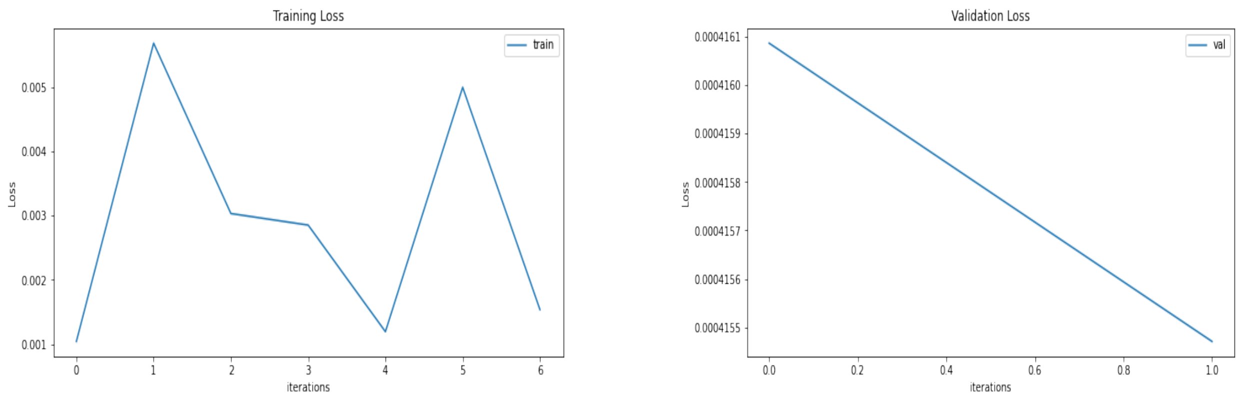


Fig. 11 Training vs. Validation Loss for AutoEncoders network pre trained with Resnet-18

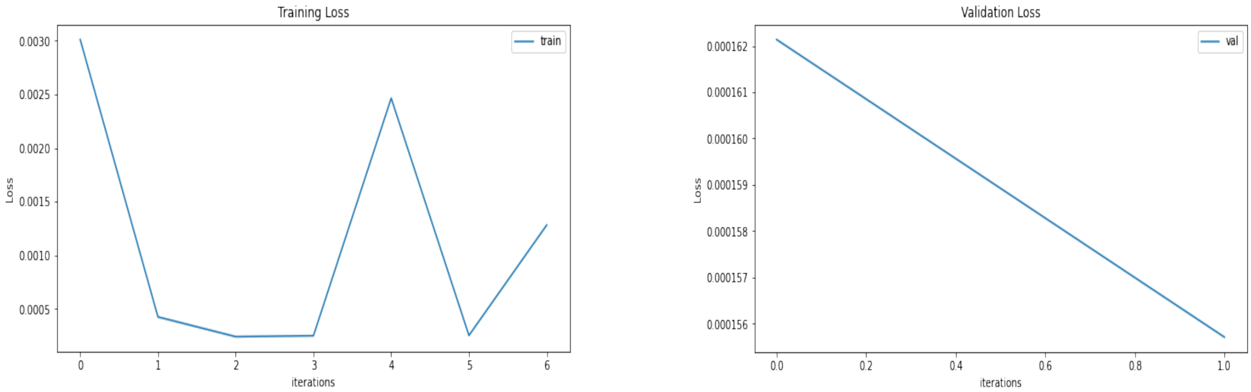


Fig. 12 Training vs. Validation Loss for Auto Encoders network pre trained with Resnet-34

After training the models on the X-Ray images the authors were able to colourize the irregularities in them. ResNet-18 is the best performing model. The basic autoencoder that the authors trained was not able to come up with any predominant results. Although ResNet-34 has more layers in it, it under-performed in comparison to ResNet-18. Lesser tarining loss is observed in coparision to other works. The future work of this research could entail contrasting different loss functions and identify the best suited measure.

Table 1. Training and Validation Losses

	Basic AutoEncoder	AutoEncoder with Resnet-18	AutoEncoder with Resnet-34	Treneska, S et. al. [48]
Training Loss (after last epoch)	0.1410	0.0010	0.0030	0.1510
Validation Loss (after last epoch)	0.1686	0.0004	0.0002	

8. Conclusion

This paper describes how to colourize X-ray pictures, which are often only available in monochrome. The main objective of this work on medical image colorization is to make it easier for medical professionals to identify and diagnose anomalies in X-ray images. It may be conceivable for colour embedded medical images to help with the correct and prompt diagnosis of a patient's ailment as well as the choice of the most appropriate therapy when this work has been refined to meet medical domain-specific practices. Since the research for this domain still needs more extensive work, this study leverages the transfer learning techniques to effectively introduce colour values to monochrome X-ray photos using ResNet structures. In order to develop a more robust framework, further research could extend optimising a loss function and investigate additional learning techniques that take the issue of the data's modality into account.

References

[1] B, S., & B, R. (2022, November 9). International Journal of Recent Technology and Engineering (IJRTE). Retrieved November 15, 2022, from https://www.ijrte.org/

[2] Farella, E. M., Malek, S., & Remondino, F. (2022). Colorizing the Past: Deep Learning for the Automatic Colorization of Historical Aerial Images. *Journal of Imaging*, *8*(10), 269..

[3] Garcea, F., Serra, A., Lamberti, F., & Morra, L. (2022). Data augmentation for medical imaging: A systematic literature review. *Computers in Biology and Medicine*, 106391.

[4] Khan, M. U. G., Gotoh, Y., & Nida, N. (2017, June 22). Medical *image colorization for better visualization and segmentation*. White Rose Research Online. Retrieved November 15, 2022, from https://eprints.whiterose.ac.uk/119356/

[5] Tariq, A., Gill, A. Y., & Hussain, H. K. (2023). Evaluating the Potential of Artificial Intelligence in Orthopedic Surgery for Value-based Healthcare. *International Journal of Multidisciplinary Sciences and Arts*, *2*(1), 27-35.

[6] Liang, Y., Lee , D., Li1 , Y., & Shin, B.-S. (2021, January 18). *Unpaired medical image colorization using generative ... - springer*. Unpaired medical image colorization using generative adversarial network. Retrieved November 15, 2022, from

https://link.springer.com/content/pdf/10.1007/s11042-020-10468-6.pdf

[7] R. Ribani and M. Marengoni, "A Survey of Transfer Learning for Convolutional Neural Networks," 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), 2019, pp. 47-57, doi: 10.1109/SIBGRAPI-T.2019.00010.

[8] D'Souza, G., Reddy, N. S., & Manjunath, K. N. (2023). Localization of lung abnormalities on chest X-rays using self-supervised equivariant attention. *Biomedical Engineering Letters*, *13*(1), 21-30.

[9] Mohan, R., Elsken, T., Zela, A., Metzen, J. H., Staffler, B., Brox, T., ... & Hutter, F. (2023). Neural Architecture Search for Dense Prediction Tasks in Computer Vision. *International Journal of Computer Vision*, 1-24.

[10] Morra, L., Piano, L., Lamberti, F., & Tommasi, T. (2020, October 19). *Bridging the gap between natural and medical images through ... - arxiv*. Bridging the gap between Natural and Medical Images through Deep Colorization. Retrieved November 15, 2022, from https://arxiv.org/pdf/2005.10589.pdf

[11] Huang, S., Jin, X., Jiang, Q., & Liu, L. (2022). Deep learning for image colorization: Current and future prospects. Engineering Applications of Artificial Intelligence, 114, 105006.

[12] Abbadi, N. K. E., & Razaq, E. S. (2020). Automatic gray images colorization based on lab color space. Indonesian Journal of Electrical Engineering and Computer Science, 18(3), 1501-1509.

[13] Wu, M., Jin, X., Jiang, Q., Lee, S. J., Liang, W., Lin, G., & Yao, S. (2021). Remote sensing image colorization using symmetrical multi-scale DCGAN in YUV color space. The Visual Computer, 37(7), 1707-1729.

[14] Li, B., Lu, Y., Pang, W., & Xu, H. (2023). Image Colorization using CycleGAN with semantic and spatial rationality. *Multimedia Tools and Applications*, 1-15.

[15] Cevallos, S., Pérez, N., Riofrío, D., Benítez, D., Moyano, R. F., & Baldeon-Calisto, M. (2022, July). A Deep Convolutional Autoencoder Architecture for Automatic Image Colorization. In 2022 IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI) (pp. 1-6). IEEE.

[16] Dong, X., Liu, C., Li, W., Hu, X., Wang, X., & Wang, Y. (2021). Self-supervised colorization towards monochrome-color camera systems using cycle CNN. IEEE Transactions on Image Processing, 30, 6609-6622.

[17] Kim, S., Jang, Y., & Kim, S. E. (2021). Image-Based TF Colorization With CNN for Direct Volume Rendering. IEEE Access, 9, 124281-124294.

[18] An, J., Kpeyiton, K. G., & Shi, Q. (2020). Grayscale images colorization with convolutional neural networks. Soft Computing, 24(7), 4751-4758.

[19] Joshi, M. R., Nkenyereye, L., Joshi, G. P., Islam, S. R., Abdullah-Al-Wadud, M., & Shrestha, S. (2020). Auto-colorization of historical images using deep convolutional neural networks. Mathematics, 8(12), 2258

[20] Mouzon, T., Pierre, F., & Berger, M. O. (2019, June). Joint cnn and variational model for fully-automatic image colorization. In International Conference on Scale Space and Variational Methods in Computer Vision (pp. 535-546). Springer, Cham.

[21] Nguyen, T., Mori, K., & Thawonmas, R. (2016). Image colorization using a deep convolutional neural network. arXiv preprint arXiv:1604.07904.

[22] Hensman, P., & Aizawa, K. (2017, November). cGAN-based manga colorization using a single training image. In 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (Vol. 3, pp. 72-77). IEEE.

[23] Mourchid, Y., Donias, M., & Berthoumieu, Y. (2021, January). Automatic Image Colorization based on Multi-Discriminators Generative Adversarial Networks. In 2020 28th European Signal Processing Conference (EUSIPCO) (pp. 1532-1536). IEEE.

[24] Chen, X., Duan, Y., Houthooft, R., Schulman, J., Sutskever, I., & Abbeel, P. (2016). Infogan: Interpretable representation learning by information maximizing generative adversarial nets. Advances in neural information processing systems, 29.

[25] Jin, X., Li, Z., Liu, K., Zou, D., Li, X., Zhu, X., ... & Liu, Q. (2021, October). Focusing on Persons: Colorizing Old Images Learning from Modern Historical Movies. In Proceedings of the 29th ACM International Conference on Multimedia (pp. 1176-1184).

[26] Farella, E. M., Malek, S., & Remondino, F. (2022). Colorizing the Past: Deep Learning for the Automatic Colorization of Historical Aerial Images. Journal of Imaging, 8(10), 269.

[27] Dias, M., Monteiro, J., Estima, J., Silva, J., & Martins, B. (2020). Semantic segmentation and colorization of grayscale aerial imagery with W-Net models. Expert systems, 37(6), e12622.

[28] Chen, S. Y., Zhang, J. Q., Zhao, Y. Y., Rosin, P. L., Lai, Y. K., & Gao, L. (2022). A review of image and video colorization: From analogies to deep learning. Visual Informatics.

[29] Huang, S., Jin, X., Jiang, Q., & Liu, L. (2022). Deep learning for image colorization: Current and future prospects. Engineering Applications of Artificial Intelligence, 114, 105006.

[30] Žeger, I., Grgic, S., Vuković, J., & Šišul, G. (2021). Grayscale image colorization methods: Overview and evaluation. IEEE Access.

[31] Nida, N., Sharif, M., Khan, M. U. G., Yasmin, M., & Fernandes, S. L. (2016). A framework for automatic colorization of medical imaging. *IIOAB J*, *7*, 202-209.

[32] Mathur, A. N., Khattar, A., & Sharma, O. (2021). 2D to 3D Medical Image Colorization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 2847-2856).

[33] Selvapriya B., Raghu B. (2019) Colorization using Desired Color for Medical Images. International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6S3

[34] Mooney, P. (2018, March 24). *Chest X-ray images (pneumonia)*. Kaggle. Retrieved September 17, 2022, from https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

[35] Liang, C. M., Li, Y. W., Liu, Y. H., Wen, P. F., & Yang, H. (2022). Segmentation and weight prediction of grape ear based on SFNet-ResNet18. Systems Science & Control Engineering, 10(1), 722-732.

[36] Tschannen, M., Bachem, O., & Lucic, M. (2018, December 12). *Recent advances in autoencoder-based representation learning*. arXiv.org. Retrieved November 17, 2022, from https://arxiv.org/abs/1812.05069

[37] Bank, D., Koenigstein, N., & Giryes, R. (2021, April 3). *Autoencoders*. arXiv.org. Retrieved November 17, 2022, from https://arxiv.org/abs/2003.05991

[38] Kamat, P., Sugandhi, R., & Kumar, S. (2021). Data-driven bearing fault detection using hybrid autoencoder-LSTM

deep learning approach. *International Journal of Modelling, Identification and Control*, *38*(1), 88-103.

[39] Jordan, J. (2018, March 19). *Chapter 14: Autoencoders*. deeplearningbook-notes. Retrieved November 17, 2022, from https://ucla-labx.github.io/deeplearningbook-notes/Ch14-Autoencoders.html

[40] *Autoencoders in Deep learning: Tutorial & use cases [2022]*. V7. (2022, October 21). Retrieved November 17, 2022, from https://www.v7labs.com/blog/autoencoders-guide

[41] Ramzan, F., Khan, M. U. G., Rehmat, A., Iqbal, S., Saba, T., Rehman, A., & Mehmood, Z. (2019, December 18). *A deep learning approach for automated diagnosis and Multi-class classification of alzheimer's disease stages using resting-state fmri and residual neural networks - Journal of Medical Systems*. SpringerLink. Retrieved November 18, 2022, from https://link.springer.com/article/10.1007/s10916-019-1475-2

[42] Sanakkayala, D. C., Varadarajan, V., Kumar, N., Karan, Soni, G., Kamat, P., ... & Kotecha, K. (2022). Explainable AI for bearing fault prognosis using deep learning techniques. *Micromachines*, *13*(9), 1471.

[43] Jin, X., Di, Y., Jiang, Q., Chu, X., Duan, Q., Yao, S., & Zhou, W. (2023). Image colorization using deep convolutional auto-encoder with multi-skip connections. Soft Computing, 27(6), 3037-3052.

[44] Hu, M., Bai, L., Fan, J., Zhao, S., & Chen, E. (2023). Vehicle color recognition based on smooth modulation neural network with multi-scale feature fusion. *Frontiers of Computer Science*, *17*(3), 173321.

[45] Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2016). Let there be color! Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics (ToG)*, *35*(4), 1-11.

[46] Mouzai, M., Mustapha, A., Bousmina, Z., Keskas, I., & Farhi, F. (2023). Xray-Net: Self-supervised pixel stretching approach to improve low-contrast medical imaging. *Computers and Electrical Engineering*, *110*, 108859..

[47] Tiwari, S., Jain, A., Sapra, V., Koundal, D., Alenezi, F., Polat, K., ... & Nour, M. (2023). A smart decision support system to diagnose arrhythymia using ensembled ConvNet and ConvNet-LSTM model. Expert Systems with Applications, 213, 118933.

[48] Treneska, S., Zdravevski, E., Pires, I. M., Lameski, P., & Gievska, S. (2022). GAN-Based Image Colorization for Self-Supervised Visual Feature Learning. *Sensors*, *22*(4), 1599. MDPI AG. Retrieved from http://dx.doi.org/10.3390/s22041599