**0510-7255 – Deep Learning**

**Final Project**

**NAS-Based Image Denoising**

Barak Mamistvalov 308001346

Roy Hachnochi 311120851

**Date:** 03.10.2021

**Table of Contents**

[1 Abstract 1](#_Toc82601278)

[2 Introduction 1](#_Toc82601279)

[3 Related Work 2](#_Toc82601280)

[3.1 Image denoising 2](#_Toc82601281)

[3.2 Neural Architecture Search (NAS) 3](#_Toc82601282)

[3.3 NAS Image Denoising 3](#_Toc82601283)

[4 Method 3](#_Toc82601284)

[5 Results 4](#_Toc82601285)

[6 Conclusion 5](#_Toc82601286)

[References 6](#_Toc82601287)

[Appendix A – Code Implementation 8](#_Toc82601288)

# Abstract

The vast work in Deep Learning (DL) has led to a leap in image denoising research. Most denoising works demand intensive use of resources and in consequence running time is not negligible. Over the last years, with the ongoing research in the field of Net Architecture Search (NAS), solutions with limited resources arise and allow fitting light models with high performance to diverse problems. Formally, our goal is to perform image denoising with a light model which has emerged from a NAS process. This is attained using the MobileNetV3 model after adjustments of the architecture through a NAS process that makes it suitable to run on limited resources and in reasonable time. This method makes it possible to train and test the model efficiently and thus can be comfortably run on mobile devices. Our architecture is based on a down-sample up-sample skeleton with the objective function of mean square error (MSE). We showcase our denoising results using the CelebA and FFHQ datasets presenting fast training time with only one A100 nvidia GPU and fast inference time on a CPU, achieving notable PSNR.

# Introduction

Image denoising is one of the most fundamental problems in image processing, and as such it has been explored quite extensively. Images are often corrupted by noise in acquisition and transmission, which usually degrades the quality of images. However, various image-related applications, such as aerospace, medical image analysis, object detection etc., generally require effective noise suppression to produce reliable results. Furthermore, denoising is often necessary as a pre-processing for other image/vision tasks, e.g. compression, segmentation and recognition. Therefore, denoising has been one of the most important and widely studied problems in image processing and computer vision.

As deep learning emerged in the past decade, many neural-network-based attempts were made to solve this task. These led to state-of-the-art (SoTA) performance in commonly used full reference distortion measures, such as Mean-Squared-Error (MSE), that quantify the discrepancy between the denoised image and its clean source ( [1], [2], [3], [4], [5], [6]). Digital images play an important role both in daily life applications such as satellite television, autonomous vehicles, airplane monitor etc. These applicants use videos, and the importance of real-time inputs is undebatable. In addition, there is a great significance that each frame in the video will be received without noise. Most works in the field of image denoising concentrate their efforts on the distortion measure and architecture choosing, and they indeed achieve low distortion with heavy models. In our work, we suggest a denoiser which is a neural network that is relatively light and has reasonable training and, more importantly, inference times. We use the research improvement in the field of Neural Architecture Search ([7], [8], [9]) to get a light model that achieves good results with low running time. The growing interest in the automation of machine learning and deep learning has inevitably led to the development of a wide variety of automated methods for neural architecture search. The choice of the network architecture has proven to be critical, and many advances in deep learning spring from its immediate improvements. In our work, we chose an architecture which is a result of a NAS process on the MobileNetV3 ( [10]). The model we used consists mainly of bottleneck blocks which reduce the resolution to small image features, and of up-sample blocks which ultimately supply the size of the input image. The bottleneck parameters are satisfied through the NAS process from [11] which was at first designated for a classification problem. In our work, we have transformed this architecture for the denoising task. We conducted only minor hyper-parameters tuning, but more tuning may achieve better results of runtime and distortion elimination. In the results section we present a model with competitive performance metrics, achieved with low inference times on a CPU as well as on a GPU, allowing the usage of the architecture on compact devices.

# Related Work

## Image denoising

Image denoising methods are in continuous research and there are many classical and learning-based algorithms attempting to tackle this problem in the past 50 years ([12], [13]). Among the classical algorithms we may find the support vector decomposition (SVD) denoising ([14], [15], [16]) which propose a computationally simple denoising algorithm using the nonlocal self-similarity. In case there is a prior on the noise distribution, Wiener filtering may be applied in the power spectrum ([17]). Wavelets are also a common classical tool for image denoising ([18], [19]) when the main aim is to modify the wavelet coefficients in the new basis, so the noise can be removed from the data.

Since the proposal of big data analysis and Graphic Processing Unit (GPU), the deep learning technique has received a great deal of attention and has been widely applied in the field of image processing. Many deep learning algorithms and architectures have emerged to cope with the image denoising task. First used by Zhou et al. ([20]) suggest a neural network with both the known shift-invariant blur function and additive noise to recover the latent clean image. Agostinelli et al. ([21]) suggest the BRDNet which is a CNN based architecture that combines two networks to increase the width of the network, and thus obtains more features in the process. Several works used GANs to deal with the problem. Linh Duy Tran et al. ([22]) tried to deal with real images noise and not necessarily synthetic one and utilize GANs to estimate the noise distribution at first and add it to the inputs of a deep neural network denoiser. Ohayon et al. ([23]) suggest sampling clean images from a distribution learned by a CGAN that its conditional input is a noised image, and the output is the clean image.

Works in the deep learning field, especially those with high performance, tend to be heavy and slow, thus designated to work offline and consume many resources, memory, and time to operate. Our contribution is an efficient learning denoiser architecture that is designated to operate fast and is a light model emerging from a NAS process.

## Neural Architecture Search (NAS)

Deep Learning has enabled remarkable progress over the previous years on a variety of tasks,

such as image recognition, speech recognition, and machine translation. One crucial aspect for this progress are novel neural architectures. The early days of the deep learning era focused on utilizing neural networks for various tasks, while mostly ignoring memory and computational efficiency. Throughout the years the importance of the network’s architecture came to knowledge in research, both for performance and both for finding the most efficient architecture to perform a given task. NAS (Neural Architecture Search) are methods to automate the process of models’ architecture design per provided latency/other constraints.

NAS methods may be divided into several types:

* Random search methods – Simply randomly search the parameter and architecture space, searching for the best architecture.
* Evolutionary methods – A network is sampled every step, and then serves as a parent to generate child networks with some mutations.
* Gradient-based methods (e.g., [8]) – A continuous relaxation is proposed as the architecture-optimization problem, allowing gradient-based optimization on a combination of operations.
* Reinforcement Learning methods (e.g., [7]) – Based on the regular RL scheme, an agent generates architectures and achieves rewards in the form of performance estimations, the agent’s policy is optimized to achieve the highest reward via RL techniques.

Earlier NAS methods (such as [7]) focused solely on accuracy and not on efficiency, producing SOTA classification models, on the expense of inference time. Hardware-aware NAS ([9]) also began emerging, attempting to achieve the best performance while satisfying hardware constraints. Our model is based on such a model and is a Gradient-based method.

## NAS Image Denoising

Our focus on this project is on the task of image-denoising. Specifically, latency efficient image denoising, achieved using a NAS process. It is important to note that our project does not actually perform NAS for the desired task, but utilizes an architecture proposed by another paper for an image classification task (which was indeed acquired using NAS) and transfers it to the image denoising task. The main focus is the time efficiency of the acquired model. We therefore present some papers which did perform NAS for efficient image denoising.

TODO: continue:

<https://arxiv.org/pdf/2104.02525.pdf>

<https://arxiv.org/pdf/1909.08228.pdf>

<https://arxiv.org/pdf/2004.08870.pdf>

# Method

We adopt the architecture proposed by [11], which performed a NAS process on MobileNetV3 and found an efficient and light architecture fit for the image classification task. Their architecture is composed of several bottleneck blocks with different parameters and may be described as follows:

TODO: add diagram of architecture

The paper shows superior top-1 accuracy for small latencies, as presented:

Chart, scatter chart

Description automatically generated

Our goal is to use this architecture as a base-block for our image denoising model, retrain it, and achieve good denoising with short latency. It is well known that a good image-denoising scheme is composed of an encoder-decoder structure (references?). Consequently, we use the classifier structure proposed above as an encoder, while eliminating the final classification layers, and add are own deconvolution layers as a decoder. Our hope is that the good accuracy achieved by the original papers classifying architecture will result in a fitting encoder for our image-denoising network, since it was previously shown by a wide range of papers (references?) that good classification is achieved by an efficient encoding of the image features. We then use these features as the input to the deconvolution-based decoder, wishing to maintain the short latency achieved by the encoding architecture and use it for the image denoising task. Our architecture may be ultimately described as follows:

TODO: add diagram of architecture

TODO: add an explanation of our training process?

# Results

The model is trained using only a single GPU (Nvidia A100-SXM4-40GB) with epoch time of about 3 min when trained with input size of 32x32x3 and tens of dozens of images in the dataset. In addition, we got MSE of \_\_\_\_\_.

# Conclusion

# References

|  |  |
| --- | --- |
| [1] | R. Y. a. L. G. Rini Thakur, "State-of-art analysis of image denoising methods using convolutional neural networks," *IET Image Processing,* vol. 13, no. 8, 2019. |
| [2] | C. T. e. al., "Deep learning on image denoising: an overview," *Neural Networks,* vol. 131, pp. 251-275, 2020. |
| [3] | F. T. e. al., "Variational inference for computational imaging inverse problems," *arXiv preprint arXiv:1904.06264,* 2020. |
| [4] | M. E. a. P. M. Gregory Vaksman, "Lidia: lightweight learned image denoising with instance adaptation," *IEEE Conference on Computer Vision and Pattern Recognition Workshops,* 2020. |
| [5] | E. M. L. A. G. D. Jay Whang, "Approximate probabilistic inference with composed flows," *arXiv preprint arXiv:2002.11743,* 2020. |
| [6] | D. Y. e. al., "Diversity-sensitive conditional generative adversarial networks," *Proceedings of the International Conference on Learning Representations,* 2019. |
| [7] | B. a. L. Q. V. Zoph, " Neural architecture search with reinforcement learning," *arXiv preprint arXiv:1611.01578,* 2016. |
| [8] | H. S. K. a. Y. Y. Liu, "Darts: Differentiable architecture search," *arXiv preprint arXiv:1806.09055,* 2018. |
| [9] | H. Z. L. a. H. S. Cai, "Proxylessnas: Direct neural architecture search on target task and hardware," *arXiv preprint arXiv:1812.00332,* 2018. |
| [10] | A. P. R. A. H. L. Q. V. S. M. C. B. W. W. C. L. T. M. C. G. V. V. a. Z. Y. Howard, "Searching for mobilenetv3," *ICCV,* p. 1314–1324, 2019. |
| [11] | Y. A. A. N. a. L. Z.-M. Niv Nayman, "HardCoRe-NAS: Hard Constrained diffeRentiable Neural Architecture Search," *arXiv preprint arXiv:2102.11646,* 2021. |
| [12] | B. R., "Adaptive nonlinear filters for simultaneous removal of different kinds of noise in images," *IEEE Transactions on Circuits and Systems,* vol. 34, no. 11, pp. 1275-1291, 1987. |
| [13] | J. Z. L. Z. W. Z. D. &. F. X. Xu, "Patch group based nonlocal self-similarity prior learning for image denoising," *In Proceedings of the IEEE international conference on computer vision,* p. 244–252, 2015. |
| [14] | J. M. E. a. A. W. Orchard, "Efficient nonlocal-means denoising using the SVD," *2008 15th ieee international conference on image processing,* 2008. |
| [15] | Q. e. a. Guo, "An efficient SVD-based method for image denoising," *IEEE transactions on Circuits and Systems for Video Technology,* vol. 26, no. 5, pp. 868-880, 2015. |
| [16] | M. a. A. L. Lebrun, "An implementation and detailed analysis of the K-SVD image denoising algorithm," *Image Processing On Line 2,* pp. 96-133, 2012. |
| [17] | S. a. T. S. Suhaila, "Power spectrum estimation method for image denoising by frequency domain Wiener filter," *ICCAE,* vol. 3, 2010. |
| [18] | S. D. a. D. D. D. Ruikar, "Wavelet based image denoising technique," *IJACSA,* vol. 2, no. 3, 2011. |
| [19] | F. T. B. a. M. U. Luisier, "A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding," *IEEE Transactions on image processing,* vol. 16, no. 3, pp. 593-606, 2007. |
| [20] | Y. C. R. &. J. B. Zhou, "A novel approach to image restoration based on a neural network," *Proceedings of the international conference on neural networks, San Diego, California,* 1987. |
| [21] | F. M. R. A. a. H. L. Agostinelli, "Adaptive multi-column deep neural networks with application to robust image denoising," *Advances in neural information processing systems,* 2013. |
| [22] | S. M. N. M. A. Linh Duy Tran, "GAN-based Noise Model for Denoising Real Images," *ACCV,* 2020. |
| [23] | e. a. Ohayon Guy, "High perceptual quality image denoising with a posterior sampling cgan," *arXiv preprint arXiv:2103.04192,* 2021. |

# Appendix A – Code Implementation

Our work may be found fully in . We based our model on <https://github.com/Alibaba-MIIL/HardCoReNAS>