

Enhancing Image Resolution through Generative Adversarial Deep Learning Networks for Superior Quality Images

Repana Somasekhar

Department of Computer Science Sri Venkateswara University, Tirupati

Abstract— Face Recognition (FR) is crucial in computer vision, particularly for identifying faces across distributed cameras. However, efficiency decreases due to limited references and differences between faces in Enrollment Domain (ED) and Operational Domain (OD). This paper proposes a comprehensive solution addressing various FR challenges. It utilizes deep learning models fed with domain-specific dictionaries and employs face illumination transfer techniques to tackle illumination issues. Additionally, it trains Super-Resolution Generative Adversarial Network (SRGAN) and Deblur Generative Adversarial Network (DeblurGAN) to address low-resolution and blurriness problems, respectively, thereby enhancing image super resolution using Generative Adversarial Networks.

I. INTRODUCTION

Face Recognition (FR) is a significant aspect of computer vision, used for identifying faces across networked cameras. This paper addresses challenges in FR, particularly with Single Sample Per Person (SSPP) and differences between faces in Enrollment Domain (ED) and Operational Domain (OD). It proposes a comprehensive solution involving deep learning models fed with domain-specific dictionaries and employing face illumination transfer techniques to overcome illumination issues. Additionally, it trains Super-Resolution Generative Adversarial Network (SRGAN) and Deblur Generative Adversarial Network (DeblurGAN) to address low-resolution and blurriness problems, respectively.

1.1 Face Recognition

Face Recognition (FR) is a significant field in computer vision, widely used in surveillance, law enforcement, and video monitoring. FR faces challenges due to limited references. Various techniques address this, including traditional methods like Support Vector Machine (SVM) and Principal Component Analysis (PCA) for feature extraction. Dimensional recognition, utilizing 3D sensors, enhances efficiency by capturing detailed face shapes. Skin texture analysis and thermal cameras offer additional insights, improving recognition accuracy. Despite advancements like deep learning, challenges persist, especially with limited data scenarios.

1.2 Image Super Resolution

Super resolution enhances image quality by combining multiple low-resolution images to produce a higher-resolution result. This process overcomes sensor limitations without additional hardware. The goal is to increase image resolution, measured by frequency content. Obtaining high-resolution images can be costly and challenging. Super resolution algorithms face various challenges alongside their main task. In addition to being able to compute values for all the super resolution image pixels intensities given the low-resolution image pixel intensities, a super-resolution system must also be able to handle:

- **Image registration** – small image displacements are crucial for beating the sampling limit of the original camera, but the exact mappings between these images are unknown. To achieve an accurate super-resolution result, they need to be found as accurately as possible.
- **Lighting variation** – when the images are aligned geometrically, there may still be significant photometric variation, because of different lighting levels or camera exposure settings when the images were captured.

II. LITERATURE SURVEY

2.1 Domain-Specific Face Synthesis for Video Face Recognition from A Single Sample Per Person.

F. Mokhayeri, E. Granger (2019).

The article discusses an algorithm for domain-specific face synthesis (DSFS) using intra-class variation information from operational domain (OD). It selects a compact set of faces from unknown persons in OD using affinity propagation clustering

during camera calibration. These variations are then projected onto reference images using an image-based face relighting technique. Experimental results from Chokeypoint and COX-S2V datasets show that augmenting reference gallery sets with DSFS improves accuracy significantly, with only a moderate increase in computational complexity.

2.2 Deblurgan-Blind Motion Deblurring Using Conditional Adversarial Networks.

O. Kupyn, V. Budzan (2018).

In this article, we present DeblurGAN, an end-to-end learned method for motion deblurring. The learning is based on a conditional GAN and the content loss. DeblurGAN achieves state-of-the-art performance both in the structural similarity measure and visual appearance. The quality of the deblurring model is also evaluated in a novel way on a real-world problem are object detection on de-blurred images. The method is 5 times faster than the closest competitor DeepDeblur. We also introduce a novel method for generating synthetic motion blurred images from sharp ones, allowing realistic dataset augmentation. Experimental results obtain in the better performance, when compared to the other system.

2.3 Deep Convolutional Neural Network Used in Single Sample Per Person Face Recognition.

J. Zeng, X. Zhao, J. Gan, C. Mai (2018)

In this article, the paper proposes a scheme combined traditional and deep learning (TDL) method to process the task. First, it proposes an expanding sample method based on traditional approach. Compared with other expanding sample methods, the method can be used easily and conveniently. Second, it uses transfer learning and introduces a well-trained deep convolutional neural network (DCNN) model and then selects some expanding samples to fine-tune the DCNN model. Third, the fine-tuned model is used to implement experiment. Experimental results on AR face database, Extend Yale B face database, FERET face database, and LFW database demonstrate that TDL achieves the state-of-the-art performance in SSPP FR.

2.4 Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.

C. Ledig, L. Theis, F. Huszar (2017).

The article introduces SRGAN, a GAN for image super-resolution (SR), capable of generating photo-realistic images at 4x upscaling. It employs a perceptual loss function, combining adversarial and content losses, to achieve natural image generation. The deep residual network can recover realistic textures from heavily downsampled images, as evidenced by extensive MOS tests, indicating significant perceptual quality improvements compared to state-of-the-art methods.

2.5 Dynamic Ensembles of Exemplar-SVMS for Still-To-Video Face Recognition.

S. Bashbaghi, E. Granger (2017)

In this article, the paper demonstrated an efficient multi-classifier system (MCS) is proposed for accurate still-to-video FR based on multiple face representations and domain adaptation (DA). Specific ensemble of exemplar-SVM (e-SVM) classifiers is thereby designed to improve robustness to intra-class variations. Ensemble is used to model the single reference still, where multiple face descriptors and random feature subspaces allow to generate a diverse pool of patch-wise classifiers. This paper also investigates the impact of using different training schemes for DA, as well as, the validation set of non-target faces extracted from stills and video trajectories of unknown individuals in the operational domain. The performance of the proposed system was validated using videos from the COX-S2V and Chokeypoint datasets. Results indicate that the proposed system can surpass state-of-the-art accuracy, yet with a significantly lower computational complexity.

2.6 Dynamic Dictionary Optimization for Sparse-Representation-Based Face Classification Using Local Difference Images.

C. Shao, X. Song (2017)

The article introduces a sparse-representation-based face-classification algorithm that optimizes dictionaries dynamically using synthesized faces. It augments the dictionary with virtual faces to enhance representation capacity and uses dictionary optimization to reduce redundancy and improve accuracy. Experimental results on AR and FERRET face datasets show superior accuracy, particularly for small-sample-size problems.

2.7 Face Recognition Using Deep Multi-Pose Representations.

W. Abdalmageed, Y. Wu, S. Rawls (2016).

In this paper, we introduce our method and system for face recognition using multiple pose-aware deep learning models. In our representation, a face image is processed by several pose-specific deep convolutional neural network (CNN) models to generate multiple pose-specific features. 3D rendering is used to generate multiple face poses from the input image. Sensitivity of the recognition system to pose variations is reduced since we use an ensemble of pose-specific CNN features. The paper also presents extensive experimental results on the effect of landmark detection, CNN layer selection and pose model selection on the performance of the recognition pipeline. Our novel representation achieves better results in the state-of-the-art methods.

2.8 Facenet-A Unified Embedding for Face Recognition and Clustering.

F. Schroff, D. Kalenichenko (2015)

In this paper, we present a system, called FaceNet, that directly learns a mapping from face images to a compact Euclidean space where distances directly correspond to a measure of face similarity. Once this space has been produced, tasks such as face recognition, verification and clustering can be easily implemented using standard techniques with FaceNet embeddings as feature vectors. Our method uses a deep convolutional network trained to directly optimize the embedding itself, rather than an intermediate bottleneck layer as in previous approaches. To train, we use triplets of roughly aligned matching / non-matching face patches generated using a novel online triplet mining method. We achieve state-of-the-art face recognition performance using only 128 bytes per face and then widely used Labelled Faces in the Wild (LFW) dataset, our system achieves a new record accuracy of 99.63%.

2.9 Deep Face Recognition.

Parikh, A. Vedaldi (2015)

In this article, the Paper demonstrated the face recognition from either a single photograph or from a set of faces tracked in a video. It can be analysed the two factors (i) end to end learning for the task using a convolutional neural network (CNN), and (ii) the availability of very large-scale training datasets. We make two contributions: first, we show how a very large-scale dataset (2.6M images, over 2.6K people) can be assembled by a combination of automation and human in the loop, and discuss the trade-off between data purity and time. Second, we traverse through the complexities of deep network training and face recognition to present methods and procedures to achieve comparable state of the art results on the standard LFW and YTF face benchmarks.

2.10 A Benchmark and Comparative Study of Video-Based Face Recognition on COX Face Database.

Z. Huang, S. Shan, R. Wang (2015).

In this article, the paper implemented an contributes a benchmarking and comparative study based on a newly collected still/video face database, named COX¹ Face DB. Specifically, we make three contributions. First, we collect and release a largescale still/video face database to simulate video surveillance with three different video-based face recognition scenarios (V2S, S2V, and V2V). Second, for benchmarking the three scenarios designed on our database, we review and experimentally compare a number of existing set-based methods. Third, we further propose a novel Point-to-Set Correlation Learning (PSCL) method, and experimentally show that it can be used as a promising baseline method for V2S/S2V face recognition on COX Face DB. Extensive experimental results clearly demonstrate that video-based face recognition needs more efforts, and our COX Face DB is a good benchmark database for evaluation.

Problem Definition: Face Recognition (FR) in computer vision is crucial for identifying faces across distributed cameras. The problem is divided into recognition with multiple samples per person and Single Sample Per Person (SSPP). Efficiency decreases due to limited references, especially SSPP, and differences between faces in Operational Domain (OD) and Enrollment Domain (ED). Existing methods employ pre-trained Convolutional Neural Networks (CNN) for recognition, but Super Resolution GAN is an advanced technique in deep learning to deblur images and overcome blurriness.

Drawbacks:

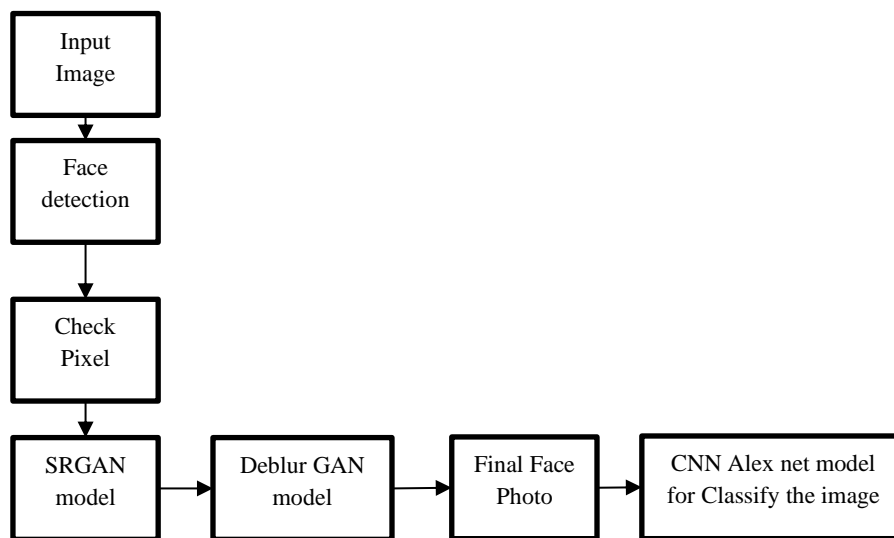
- Resize of image or not enough to apply.

- Less loss of GAN function.
- Inefficiency.
- Less Performance.

III. PROPOSED WORK

Face Recognition (FR) in computer vision is crucial for identifying faces across distributed cameras. Efficiency decreases due to limited references, especially Single Sample Per Person (SSPP), and differences between faces in Operational Domain (OD) and Enrollment Domain (ED). This paper proposes a method addressing these issues, utilizing deep learning models fed by a design domain dictionary. Face illumination transfer techniques, along with Super-Resolution Generative Adversarial Network (SRGAN) and Deblur Generative Adversarial Network (DeblurGAN), overcome illumination, low-resolution, and blurriness problems. SRGAN upscales low-resolution images, while DeblurGAN sharpens blurred images. Deep learning then identifies faces accurately and performs with high performance.

IV. IMPLEMENTATION



4.1 Datasets

All facial recognition and detection systems require the use face datasets for training and testing purposes. In particular, the accuracy of CNNs is highly dependent on large training datasets. Here, a lot of data's can be obtained to detect the face. An input image is taken as face recognition datasets. This face dataset is limited, and thus one area of improvement could include the creation of a largescale annotated dataset containing a broad range of applications.

4.2 Face Detection

Face detection is a fundamental step in face recognition and verification. Face detection methods as applied to facial recognition and verification. The greatest obstacle faced by face detection algorithms was the ability to achieve high accuracy. Consequently, their usability in real life applications was limited. So, the significant progress has since been made by system due to the development of powerful feature extraction techniques including Histograms of oriented Gradients (HoGs), Local Binary Patterns (LBPs) etc. Face detection classify the feature of the image in resolution, illumination, pose, expression, and color. In face detection process, pre-processing techniques is used to convert the image into pixel size for further process the system in efficient way.

4.3 Feature Extraction

Feature extraction usually occurs immediately after face detection and can be considered as one of the most important stages in face recognition systems, as their effectiveness is dependent upon the quality of the extracted features. Face recognition differs to object recognition in that it involves alignment before extraction. This is reflected in the differences between CNNs

used for face recognition and those used for object recognition. An increase in data availability has resulted in development of learning-based methods as opposed to engineered features due to their inherent ability to discover and optimize features specific to a task.

4.4 SRGAN Model

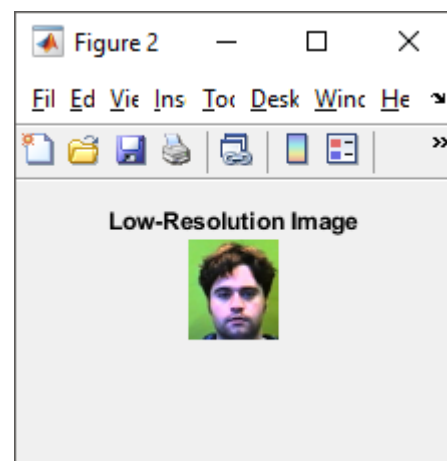
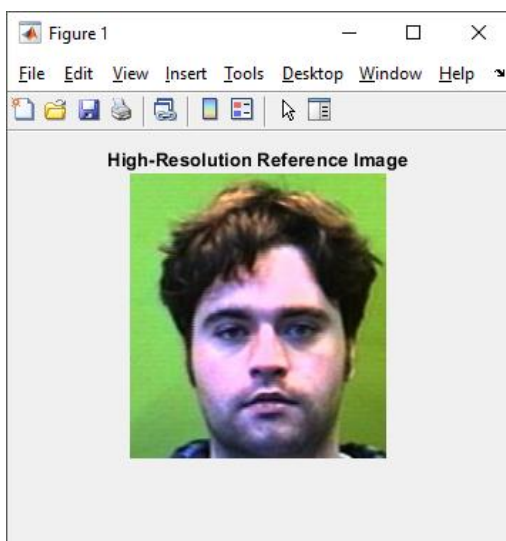
GANs are deep neural network architectures and consist of two main networks (the Generator network and the Discriminator network). The aim of GANs is to generate new data that matches the training data distribution. It is like a game in which the Generator tries to generate some data from the distribution of probability and the Discriminator serves as a judge. Discriminator determines either the input comes from a true training dataset or fake generated data. The generator attempts to optimize data in order to match true training data. The discriminator guides the generator to generate realistic data. Discriminator and generator both learn concurrently and once the generator has been trained, it has enough knowledge of the distribution of training samples. Now SRGAN is used to produce higher resolution images and we will train it and use it to overcome the problem of low-resolution faces that come from the operational domain. Processing the High-Resolution (HR) face photos for down-sampling Low-Resolution (LR) face photos. Now, we have HR and LR face photos for the training dataset. Passing LR face images through Generator which up-samples and gives Super Resolution face images. Using a discriminator to distinguish the HR face images and back-propagate the GAN loss to train the discriminator and the generator. SRGAN after training can take a low-resolution image and generate a high-resolution image. We use SRGAN after training in the operational domain. Check if the face photo less than size 96-96 (low-Resolution image). Then, the SRGAN takes the face photo and generates a high-Resolution Face photo. But the transfer image from low to high resolution, one of the drawbacks will be obtained. To overcome the drawback, Deblur GAN model is used.

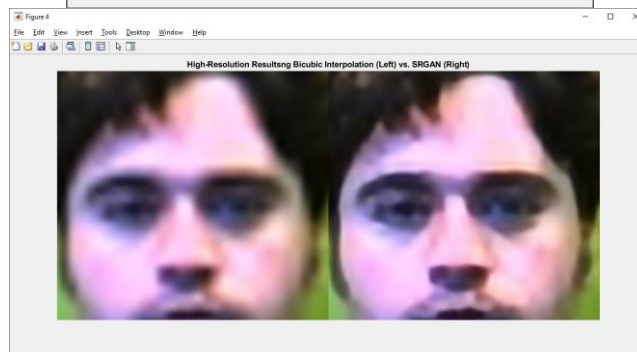
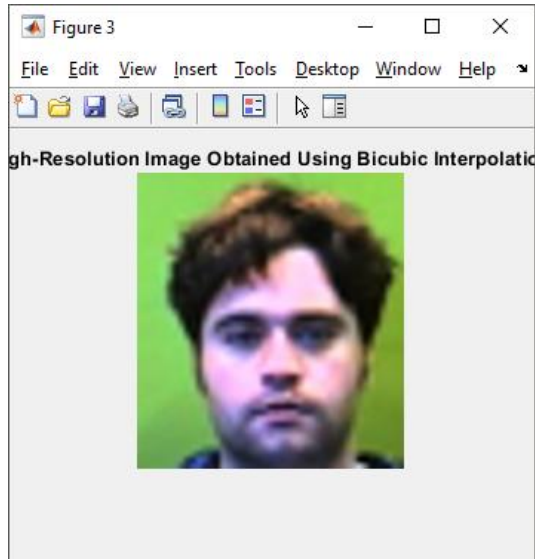
4.5 Deblur Generative Adversarial Network (Deblur-GAN)

DeblurGAN architecture consists of the network of generators that inputs the blurred image and generates a sharp image and the discriminating network to decide whether an input image is created artificially. We will train DeblurGAN to work on blurred face images. DeblurGAN after training can take a blur face image and generate a sharp face image. We use DeblurGAN after training in the operational domain. If the image that comes from the operational domain is blurred, then DeblurGAN takes the image to generate a sharp image to overcome the problem of the blurred face image.

4.6 Classification

Classification process is used in the output layer of the system. Here, CNN Alex net model is used for classify the image in face recognition system. CNN-Alexnet model is used to solving the problem of multi-classes classification. Alexnet consists of the layers of convolution and the max-layer and finished with two layers completely connected. The loss was measured using a SoftMax classifier, using multinomial logistic loss. Finally, a predicted output is obtained in this layer to classify the output images and display it.



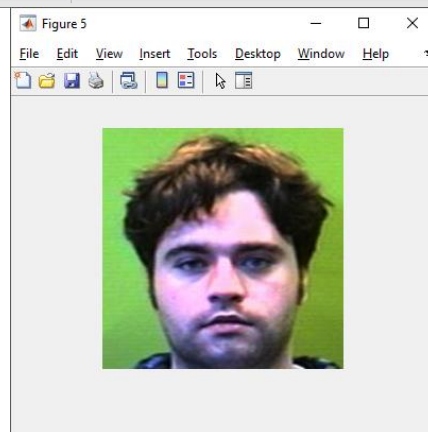


Deep Learning Network Analyzer

net
Analysis date: 12-Aug-2021 18:59:49

25 layers 0 warnings 0 errors

Name	Type	Activations	Learnables
1 data 227x227x3 images with 'zerocenter' n...	Image Input	227x227x3	-
2 conv1 96 11x11x3 convolutions with stride 4...	Convolution	55x55x96	Weights 11x11x3x96 Bias 1x1x96
3 relu1 ReLU	ReLU	55x55x96	-
4 norm1 cross channel normalization with 5 ch...	Cross Channel Nor...	55x55x96	-
5 pool1 3x3 max pooling with stride [2 2] and ...	Max Pooling	27x27x96	-
6 conv2 2 groups of 128 5x5x48 convolutions ...	Grouped Convolution	27x27x256	Weigh... 5x5x48x128... Bias 1x1x128x2
7 relu2 ReLU	ReLU	27x27x256	-
8 norm2 cross channel normalization with 5 ch...	Cross Channel Nor...	27x27x256	-
9 pool2 3x3 max pooling with stride [2 2] and ...	Max Pooling	13x13x256	-
10 conv3 256 3x3x256 convolutions with stride [...]	Convolution	13x13x384	Weights 3x3x256x384 Bias 1x1x384
11 relu3 ReLU	ReLU	13x13x384	-
12 conv4 Grouped Convolution	Grouped Convolution	13x13x384	Weigh... 3x3x192x192...



label = categorical ajflem8.

V. CONCLUSION

Face recognition is crucial and challenging, especially with Single Sample Per Person (SSPP). Our method addresses limited references and variations in illumination, pose, blurriness, and low-resolution images. We employ 3D Face Reconstruction for pose, SRGAN for low-resolution, and DeblurGAN for blurriness. Additionally, we maintain constant illumination across domains. Our approach achieves high accuracy compared to SSPP techniques and outperforms Traditional and Deep Learning (TDL) methods.

REFERENCES

- [1] Zhang, L.; Wu, X. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Trans. Image Process.* 2006, 15, 2226–2238.
- [2] Jung, S.W.; Kim, T.H.; Ko, S.J. A novel multiple image deblurring technique using fuzzy projection onto convex sets. *IEEE Signal Process. Lett.* 2009, 16, 192–195.
- [3] Nayak, R.; Harshavardhan, S.; Patra, D. Morphology based iterative back-projection for super-resolution reconstruction of image. In *Proceedings of the 2014 2nd International Conference on Emerging Technology Trends in Electronics, Communication and Networking*, Surat, India, 26–27 December 2014; pp. 1–6.
- [4] Sun, D.; Gao, Q.; Lu, Y.; Huang, Z.; Li, T. A novel image denoising algorithm using linear Bayesian MAP estimation based on sparse representation. *Signal Process.* 2014, 100, 132–145. [CrossRef]
- [5] Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution via sparse representation. *IEEE Process. IEEE Trans. Image Process.* 2010, 19, 2861–2873. [CrossRef] [PubMed]
- [6] Tang, S.; Xiao, L.; Liu, P. Single image super-resolution method via refined local learning. *J. Shanghai Jiaotong Univ. (Sci.)* 2015, 20, 26–31. [CrossRef]
- [7] He, L.; Qi, H.; Zaretzki, R. Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, ON, USA, 23–28 June 2013; pp. 345–352.
- [8] Peleg, T.; Elad, M. A statistical prediction model based on sparse representations for single image super-resolution. *IEEE Trans. Image Process.* 2014, 23, 2569–2582.
- [9] Toulson, Rob. (2008). *Advanced Rapid Prototyping in Small Research Projects with Matlab/Simulink.*