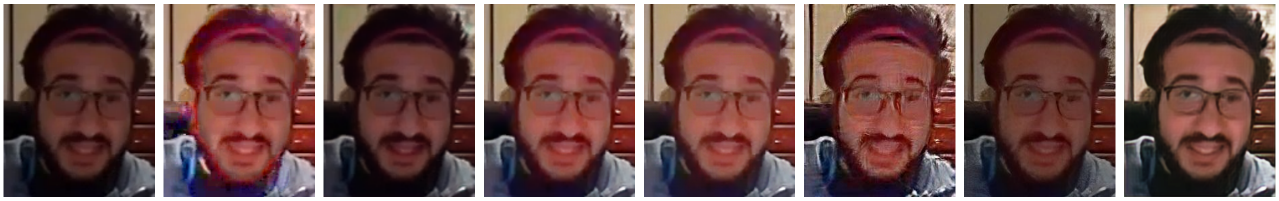
Network Architecture Search for Face Enhancement

Rajeev Yasarla[1], Hamid Reza Vaezi Joze[2], and Vishal M. Patel[1]

[1]Department of Electrical and Computer Engineering, Johns Hopkins University
[2]Microsoft
{ryasarl1,vpatel36}@jhu.edu, hava@microsoft.com



Degraded image   Super-FAN [1]   Shen *et al.* [2]   UMSN [3]   DeblurGANv2 [4]   DFDNet [5]   HiFaceGAN [6]   NASFE (ours)

Fig. 1: Sample results on the face with multiple degradations like blur, noise and low-light conditions. Restoration methods such as [1]–[6] fail to reconstruct a high quality clean face image. In constrast, the proposed NASFE network produces a high quality face image.

***Abstract*—Various factors such as ambient lighting conditions, noise, motion blur, etc. affect the quality of captured face images. Poor quality face images often reduce the performance of face analysis and recognition systems. Hence, it is important to enhance the quality of face images collected in such conditions. We present a multi-task face restoration network, called Network Architecture Search for Face Enhancement (NASFE), which can enhance poor quality face images containing multiple degradations (noise+blur+low-light). During training, NASFE uses clean face images of a person present in the degraded image to extract the identity information in terms of features for restoring the image. Furthermore, the network is guided by an identity-loss so that the identity information is maintained in the restored image. Additionally, we propose a network architecture search-based fusion network in NASFE which fuses the task-specific features that are extracted using the task-specific encoders. We introduce FFT-op and deveiling operators in the fusion network to efficiently fuse the task-specific features. Comprehensive experiments on synthetic and real images demonstrate that the proposed method outperforms many recent state-of-the-art face restoration and enhancement methods in terms of quantitative and visual performance.**

I. Introduction

In the era of COVID-19, the use of video communication tools such as Zoom, Skype, Webex, MS Teams, Google Meet, etc. has increased drastically. In many cases, images/videos captured by these video conferencing tools are of poor quality due to low-light ambient conditions, noise, motion artifacts, etc. Fig. 1 shows an example of such an image taken during a video conference. Hence, it is important to enhance the quality of face images collected in such conditions. Furthermore, restoration of degraded face images is an important problem in many applications such as human-computer interaction (HCI), biometrics, authentication, surveillance, and face recognition.

Existing face image restoration and enhancement methods are designed to address only a single type of degradation such as blur, noise, or low-light. However, in practice face images might have been collected in the presence of multiple degradations (i.e. noise + blur+ low-light). Hence, it is important to enhance the quality of face images collected in such conditions. In this paper, we address the problem of restoring a single face image degraded by multiple degradations (noise+blur+low-light). In particular, we develop a multi-task image restoration framework where a single network is able to remove the effects of low-light conditions, noise, and blur simultaneously.

The proposed multi-task face restoration problem can be considered as a many-to-one feature mapping problem, i.e extracting task-specific features (i.e. noisy, blur, and low-light enhancement features) and fusing them to get features corresponding to the clean image. The fused features can then be used by a decoder to restore the face image. One can clearly see the importance of fusion in this framework. Rather than naively using Res2Blocks [7] or convolutional blocks in the fusion network, one can learn architecture for fusion which may lead to better restoration. To this end, we propose a neural architecture search-based approach [8], [9] for learning the fusion network architecture. Additionally, we introduce FFT-op and deveiling operators in order to process the task-specific features efficiently where these operators address image formation formulation of these multiple degradations. FFT-op is motivated by Weiner deconvolution and helps in learning the weights to efficiently fuse the task-specific features and remove the effect of blur from the features. Deveiling operator is introduced to learn the weights in order to enhance low-light conditions efficiently. Furthermore, we use a classification

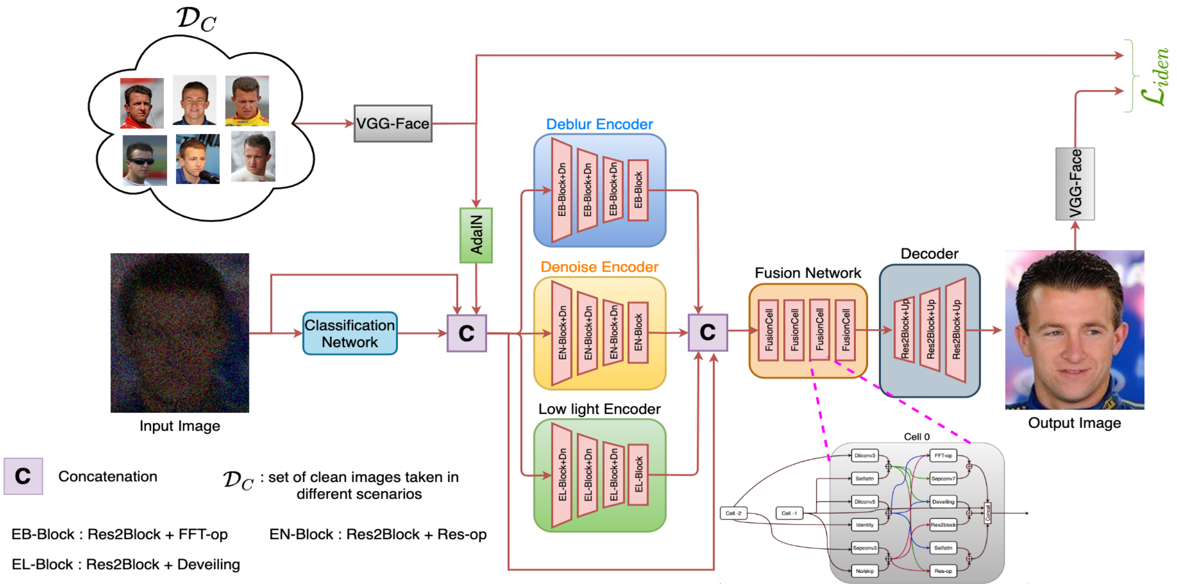


Fig. 2: An overview of the proposed NASFE face image restoration network.

network to classify the input degraded image into different classes that give information about the degradations that are present in the input face image. This class specific information is used as a prior information in the fusion network for fusing the task-specific features. Additionally, in many cases we have access to clean images corresponding to the same person in the degraded image. Hence we propose a novel way of using these clean images, which may be taken in different scenarios at different times, and extract the identity information using VGGFace features [10] and use it as prior to network along with degraded image. We also use these extracted VGGFace features and propose a novel identity-loss, $\mathcal{L}_{iden}$, for training the network. Fig. 1 shows sample results from the proposed Network Architecture Search for Face Enhancement (NASFE) method, where one can see that NASFE is able to provide better restoration results as compared to the state-of-the-art face restoration methods.

To summarize, this paper makes the following contributions:

- We propose a way of extracting the identity information from different clean images of a person present in the degraded image to restore the face image.
- We propose a novel loss, called identity-loss ($\mathcal{L}_{iden}$), which uses the aforementioned identity information to train the NASFE network.
- We propose a neural architecture search-based method for designing the fusion network.

II. Proposed Method

An observed image y, with multiple degradations, can be modeled as follows,

$$y = r \odot (k * x) + n, \tag{1}$$

where x is the clean image, r is the irradiance map, k is the blur kernel, and n is the additive noise. Here, $*$ and $\odot$ denote convolution and element-wise multiplication operations, respectively. To address this multiple degradations problem, we develop a multi-stream network called NASFE which consists of three task-specific encoders, a fusion network, and a decoder as shown in Fig. 2. The deblur encoder ($E_B(.)$), denoise encoder ($E_N(.)$) and low-light encoder ($E_L(.)$) are trained to address the corresponding tasks of deblurring, denoising, and low-light enhancement, respectively. These encoders have different building blocks that help in addressing their respective task as shown in Fig. 2. These encoders are used to extract task-specific features which are then fused using a network architecture search (NAS) based fusion block. Finally, the fused features are passed through the decoder network to restore the face image. Additionally, with help of a classification network, we determine what degradations are present in the input image, and use them as a prior information to the task-specific encoders and the fusion network. Furthermore, to improve the quality and preserve the identity in the restored face image, we extract identity information from a set of clean images corresponding to the same identity present in the degraded image in the form of VGGFace features. We denote them as the identity information (I_{iden}), and pass them as input along with the degraded image to the NASFE network to restore the face image. Besides using this identity information, we construct an identity loss $\mathcal{L}_{iden}$ to train the NASFE network.

A. Fusion Network

As can be seen from Fig. 2, the task-specific features are concatenated and then fed into the fusion network. Naively fusing these task-specific features may not be beneficial in addressing multiple degradations. Given multiple operations like FFT operator, Res-op, deveiling operator, Res2Block, and self-attention, etc., it would require significant human efforts to come up with an architecture for the fusion network that achieves good performance in addressing this problem. In order to overcome this, we follow a NAS-based approach [9], [11]–[13] for designing the fusion network. NAS-based approaches provide an elegant way to design the network architecture given the different operations required for the task. Here in our Fusion Network, we define fusion cell as the smallest repeatable module used to construct the fusion network (see Fig. 2). In our approach, the network search space includes both network-level search (i.e. searching the connection between different fusion cells), and cell level search (i.e. exploring the structure of the fusion cell).

Fusion Cell Architecture. We adopt the cell design structure of [13] to define a fusion cell (represented as $Cell(.)$) as a directed acyclic graph consisting of B blocks. In each building block the following operators are used to create the search space and construct the cell structure,

- Dilconv3 (dilated conv3 × 3)
- Dilconv5 (dilated conv5 × 5)
- separable conv 3×3
- separable conv 5×5
- Identity or skip connection
- No or zero connection
- Selfattn (Self attenion)
- Res2Block
- **Res-op (residual operator)**
- **Deveiling operator**
- **FFT operator**

Along with the conventional convolution operators like dilated, separable convolutions, no and skip connections, Res2Block [7] and self-attention block [14], [15], we introduce Res-op, deveiling, and FFT-op (shown in Fig. 3) to efficiently process the task-specific features. These operations are based on the image formation models regarding the individual degradation like noise, blur, and low-light conditions. Note, We use the same cell search technique introduced [8], [9], to search and construct or learn the cell structure.

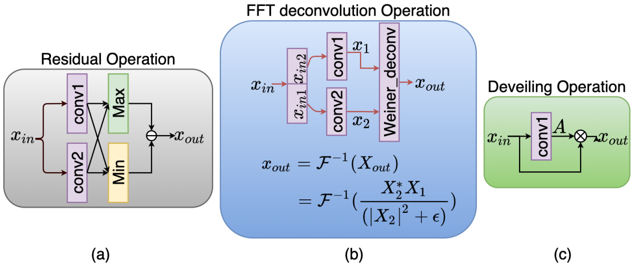

FFT deconvolution Operation
Residual Operation
x_{in}
conv1
conv2
Max
Min
x_{out}
x_{in2}
x_{in1}
x_1
x_2
Weiner_deconv
$x_{out} = \mathcal{F}^{-1}(X_{out})$
$= \mathcal{F}^{-1}(\frac{X_2^* X_1}{(|X_2|^2 + \epsilon)})$
Deveiling Operation
A
(a)
(b)
(c)

Fig. 3: (a) Residual operator, (b) FFT-operator, (c) Deveiling operator. All the convblocks (conv1 and conv2) in these operators are 3×3 convolutions.

B. Identity Information

Given the degraded image y, and a set of clean images $\mathcal{D}_C = \{C_i\}_{i=1}^n$, we compute pool3 features using the VGGFace network [10]. Note that $\mathcal{D}_C$ contains clean images of the same person present in the degraded image y. Let F^y and $\{F_i^C\}_{i=1}^n$ denote the VGGFace features corresponding to y and $\{C_i\}_{i=1}^n$, respectively. Since the clean images in $\mathcal{D}_C$ may have different style and characteristics as they may have been taken in different scenarios and times, we apply Adaptive-Instance normalization (AdaIN) before passing them as input to the network to reduce the effect of different styles in the images. AdaIN is applied as follows,

$$\bar{F}_i^C = \sigma(F^y)\left(\frac{F_i^C - \mu(F_i^C)}{\sigma(F_i^C)}\right) + \mu(F^y), \tag{2}$$

where $\sigma(.)$ and $\mu(.)$ denote standard deviation and mean, respectively. Mean of $\bar{F}_i^C$ is defined as the identity information, i.e. $I_{iden} = mean(\{\bar{F}_i^C\})$. I_{iden} is used along with y as input to the NASFE network to restore the face image.

Note that as we are extracting the identity information from the clean images, this information is much more reliable and provides stronger prior as compared to face exemplar masks [16] or semantic maps [1]–[3] extracted using the degraded images. Additionally, to preserve the identity of the subject in the restored image, we construct an identity loss $\mathcal{L}_{iden}$ to train the NASFE network.

Identity Loss $\mathcal{L}_{iden}$. Let $\hat{x}$ denote the restored face image using the NASFE network. We construct the identity loss as follows

$$\mathcal{L}_{iden} = \frac{1}{n}\sum_{i=1}^{n} \arccos(\langle F^{\hat{x}}, F_i^{\bar{C}} \rangle), \tag{3}$$

where $F^{\hat{x}}$ denotes the VGGFace features corresponding to $\hat{x}$ and n denotes the number of clean images in $\mathcal{D}_C$.

C. Overall Loss

The NASFE network is trained using a combination of the L2 loss, perceptual loss [17] and identity loss as follows

$$\mathcal{L}_{final} = \mathcal{L}_2 + \lambda_{per}\mathcal{L}_{per} + \lambda_{iden}\mathcal{L}_{iden} \tag{4}$$

where $\mathcal{L}_2 = \|\hat{x} - x\|_2^2$, $\mathcal{L}_{iden}$ denotes the identity loss defined in (3) and $\mathcal{L}_{per}$ denotes the perceptual loss [17] defined as follows

$$\mathcal{L}_{per} = \frac{1}{NHW}\sum_i \sum_j \sum_k \|F_{i,j,k}^{\hat{x}} - F_{i,j,k}^{x}\|. \tag{5}$$

Here, $F^{\hat{x}}$, F^x denote the *pool3* layer features of the VGGFace network [10] and N, H and W are the number of channels, height and width of $F^{\hat{x}}$, respectively. We set $\lambda_{per} = 0.04$ and $\lambda_{iden} = 0.003$ in our experiments.

Note that multiple clean images are required only during training. Once the network is trained, a degraded image is fed into the network and the NASFE produces an identity-preserving restored image as the output.

III. NASFE Implementation Details

Given x, we first convolve it with k to get a blurry image. Here, k can be a motion blur kernel [18], [19] or an anisotropic Gaussian blur kernel [20]. To generate a degraded image with blur+low light+noise conditions, we follow [21], [22] and convert the obtained blurry image to irradiance image L. We then multiply low light factor r to L, where $L = Mf^{-1}(k*x)$, $f(.)$ is the camera response function (CRF) function, and $M(.)$ represents the function that converts an RGB image to a Bayer image. Finally, we add realistic Photon-Gaussian noise [21], where n consists of two components: stationary noise n_c with noise variance σ_c^2 and signal dependent noise n_s with spatially varying noise variance σ_s^2.

Training dataset. We conduct our experiments using clean images from the CelebA [23] and VGGFace2 [24] face datasets.

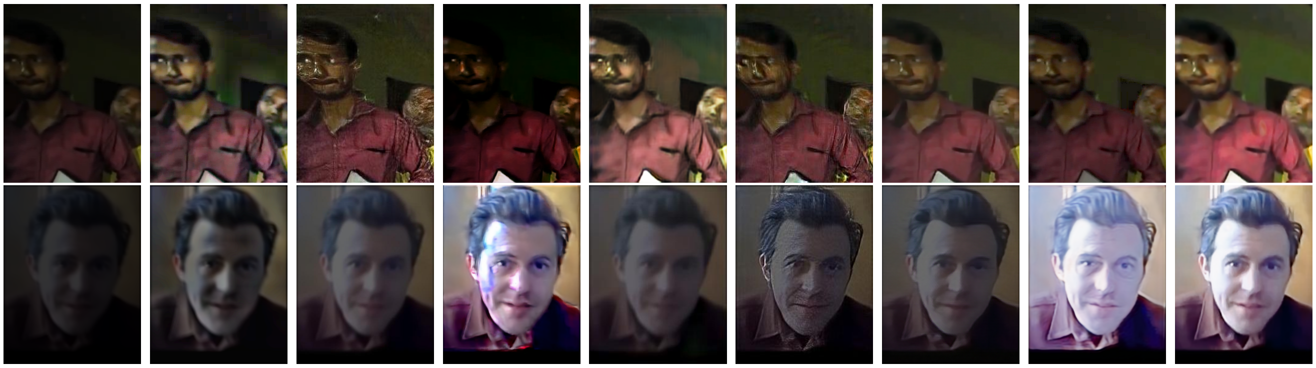


Fig. 4: Qualitative comparisons using real face images multiple degradations collected from YouTube videos. output images of Histeq are computed using Histogram equalization method.

TABLE I: PSNR|SSIM and face recognition comparisons of NASFE using *Test-BNL*. B: blur, N: noise, L: low-light. (Note all methods are retrained using degraded images containing blur, noise and low-light conditions. Histeq means Histogram equalization method)

Test-set		Metrics	B+N+L	Histeq	Shen *et al.* [2] (CVPR'18)	Super-FAN [1] (CVPR'18)	UMSN [3] (TIP'20)	DeblurGANv2 [4] (ICCV'19)	DFDNet [5] (ECCV'20)	HiFaceGAN [6] (ACMM'20)	NASFE w/o I_{iden}	NASFE (ours)
Test-BNL	CelebA	PSNR|SSIM	8.17|0.22	12.28|0.39	19.98|0.66	19.45|0.60	21.10| 0.72	21.37|0.74	21.68|0.77	21.75|0.77	22.53|0.81	**23.80**|**0.85**
		Top-1|Top-5	38.5|45.2	41.3|47.5	51.2|57.8	48.8|55.9	59.7|67.7	61.2 | 68.6	63.4|69.7	62.9|69.1	69.6|77.3	**73.4**|**81.8**
	VGGFace2	PSNR|SSIM	8.39|0.25	12.84|0.41	20.46|0.68	19.83|0.64	21.28| 0.72	21.94|0.78	21.96|0.78	22.08|0.79	22.84|0.83	**24.15**|**0.86**
		Top-1|Top-5	42.8|49.2	43.2|50.8	56.9|62.7	53.8|59.2	62.8| 69.9	65.3|72.4	67.8|73.7	67.2|72.9	72.1|80.4	**75.9**|**84.2**

TABLE II: Details of the test datasets created using CelebA [10] and VGG-Face2 [24]. M: motion kernels [18], Gaussian: Gaussian kernels [20]

Test set name	Degradation type	Details about degradation values	Number of images	
			CelebA	VGGFace2
Test-BNL	blur + noise + low-light	M: 40 kernels kernel size $[13, 27]$ G: 12 anisotropic kernels with $\sigma_b \in [1, 3]$ $r = [0.15, 0.3], \sigma_s = 0.1, \sigma_c = 0.05$	10400	10400

We randomly selected 30000 images from the training set of CelebA [23], and 30000 images from VGGFace2 [24] and generate synthetic degraded images with multiple degradations. Images in the CelebA and VGGFace2 datasets are of size 176 × 144 and 224 × 224, respectively. Given a clean face image x, we first convolve it with blur kernel k sampled from 25000 motion kernels generated using [18], [19], and 8 anisotoric Gaussian kernels [20], and then following [22], [25] we multiply them with low light factor (r) sampled uniformly from [0.05, 0.5] to obtain images with low-light conditions. Finally, we add realistic noise [21] n (where $\sigma_s \in [0.01, 0.16]$ and $\sigma_c \in [0.01, 0.06]$) to obtain the degraded image y. Based on the degradations present in y, we create class label c which is a vector of length three, $c = \{b, n, l\}$ where b, n, and l are binary numbers, *i.e* b, n, and l are one if y contains blur, noise and low-light, respectively and zero otherwise.

Test datasets. We create test datasets using randomly sampled 100 test images from the test sets of CelebA [23] and VGGFace2 [24]. Using these clean images, we create test datasets *Test-BNL* with the amounts of degradations as shown in the Table II. Additionally, we collected a real-world face image dataset with multiple degradations corresponding to 20 subjects from YouTube.

Training Details. The NASFE network is trained using $\{y_i, x_i, c_i, \mathcal{D}_C^i\}_{i=1}^N$. The classification network (CN) is trained using $\{y_i, c_i\}_{i=1}^N$. It is trained to produce a class label $\hat{c}_i$ which indicates degradation(s) present in y_i. Note CN network is combination of pretrained VGGFace [10] and three fully connected layers.

NASFE contains three encoders (deblur ($E_B(.)$), denoise ($E_N(.)$) and low-light ($E_L(.)$)), one fusion network ($Fn(.)$) and a decoder ($De(.)$) as shownn in the Fig. 2. Encoders (E_B, E_N, E_L) are initially trained to address the corresponding individual tasks of deblurring, denoising, and low-light enhancement, respectively. Given a degraded image y, we compute class $\hat{c}$ (using CN) and identity information I_{iden} and pass them as input to NASFE to compute a restored image $\hat{x}$. We set the number of blocks B to 12 in the Fusion cell of NASFE. NASFE is trained using $\mathcal{L}_{final}$ with the Adam optimizer and batch size of 40. The learning rate is set equal to 0.0005. NASFE is trained for one million iterations.

IV. Experiments and Results

We compare the performance of our method on *Test-BNL* which contains multiple degradations blur, noise, and low-light conditions. Note, we retrain [1]–[6] using degraded images that contain all degradations. Results are corresponding to this experiment are shown in Table I. As can be seen from this table, NASFE performs better by 2.1dB in PSNR and 0.07 in SSIM compared to the second-best performing method. Fig. 4 shows the qualitative results of NASFE against other methods. The outputs of other methods are blurry or contain artifacts near the eyes, nose, and mouth. On the other hand, NASFE produces clear and sharp face images. Furthermore, we can observe from Fig. 4 that outputs of other methods still contain low-light conditions, whereas NASFE produces sharp face images with enhanced lighting conditions.

We conducted face recognition experiments using *Test-BNL*, to show the significance of various face restoration methods on face recognition. Face recognition experiments are conducted using ArcFace [26] on *Test-BN*, where Top-K similar faces for the restored face image are picked from the gallery set and used to compute the accuracy. Table I show the face recognition accuracies corresponding to different methods. We can clearly see that NASFE achieves an improvement of 7% over the second-best performing method.

TABLE III: PSNR|SSIM comparisons for ablation study using *Test-BN*, *Test-BNL*. Learnable parameters are in millions.

Method	*Test-BN*		*Test-BNL*		Learnable
	CelebA	VGGFace2	CelebA	VGGFace2	Parameters
baseline network (BN)	19.72|0.70	19.88|0.72	19.95|0.65	20.42|0.69	6.00
BN-NAS	22.51|0.74	22.62|0.75	21.17|0.70	21.47|0.74	6.25
+ classification network	23.28|0.76	23.04|0.76	21.56|0.72	21.95|0.76	6.25
+ identity informaation I_{iden}	24.88|0.81	24.60|0.82	23.06|0.80	23.47|0.80	6.40
NASFE (w/ $\mathcal{L}_{mse}$ and $\mathcal{L}_{per}$)	25.10|0.84	24.92|0.84	23.35|0.83	23.76|0.83	6.40
NASFE w/ $\mathcal{L}_{final}$	25.57|0.87	25.49|0.87	23.80|0.85	24.15|0.86	6.40

V. Ablation study

We conduct ablation studies using the test-sets *Test-BN*, and *Test-BNL* to show the improvements achieved by the different components in NASFE. We start with the baseline network (BN), and define it as a combination of three encoders (E_B, E_N, and E_L), a fusion network (composed of 4 Res2Blocks [7]), and a decoder (De). As shown in Table III, BN performs very poorly due to its inability in processing task-specific features efficiently. Now, we introduce network architecture search in the fusion network by using fusion cells in-order to efficiently processing the task-specific features. The use of network architecture search results in improvement of BN-NAS by $\sim$ 2dB compared to BN. Then, we use class labels c (computed using classification network) as input to BN-NAS which increases the performance of the network by $\sim$ 0.5dB. Now we use the proposed identity information I_{iden} of the identity present in the degraded image (refer to section II-B) which further improves the performance of the network by $\sim$ 1.5dB. The resultant network corresponds to NASFE. Note that BN and BN-NAS are trained using $\mathcal{L}_{mse}$. Now we train NASFE with $\mathcal{L}_{mse}$ and $\mathcal{L}_{per}$ which further improves the performance by 0.3dB. Now we use the proposed $\mathcal{L}_{iden}$ to construct $\mathcal{L}_{final}$ and train NASFE. The proposed $\mathcal{L}_{iden}$ improves the performance of NASFE by $\sim$ 0.5dB.

VI. Conclusion

We proposed a multi-task face restoration network, called NASFE, that can enhance poor quality face images containing multiple degradations (noise+blur+low-light). NASFE makes use of the clean face images of a person present in the degraded image to extract the identity information and uses it to train the network weights. Additionally, we use network architecture search to design the fusion network in NASFE that fuses the task-specific features obtained from the encoders. Extensive experiments show that the proposed method performs significantly better than SOTA image restoration/enhancement methods on both synthetic degraded images as well as real-world images with multiple degradations (noise+blur+low-light).

References

[1] A. Bulat and G. Tzimiropoulos, "Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans," in *IEEE CVPR*, 2018.

[2] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," *IEEE CVPR*, 2018.

[3] R. Yasarla, F. Perazzi, and V. M. Patel, "Deblurring face images using uncertainty guided multi-stream semantic networks," *IEEE TIP*, vol. 29, 2020.

[4] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *IEEE ICCV*, 2019.

[5] X. Li, C. Chen, S. Zhou, X. Lin, W. Zuo, and L. Zhang, "Blind face restoration via deep multi-scale component dictionaries," in *ECCV*, 2020.

[6] L. Yang, C. Liu, P. Wang, S. Wang, P. Ren, S. Ma, and W. Gao, "Hifacegan: Face renovation via collaborative suppression and replenishment," *Proceedings of the 28th ACM International Conference on Multimedia*, 2020.

[7] S. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. H. Torr, "Res2net: A new multi-scale backbone architecture," *IEEE TPAMI*, 2019.

[8] H. Liu, K. Simonyan, O. Vinyals, C. Fernando, and K. Kavukcuoglu, "Hierarchical representations for efficient architecture search," *arXiv preprint arXiv:1711.00436*, 2017.

[9] H. Liu, K. Simonyan, and Y. Yang, "Darts: Differentiable architecture search," *arXiv preprint arXiv:1806.09055*, 2018.

[10] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[11] H. Pham, M. Y. Guan, B. Zoph, Q. V. Le, and J. Dean, "Efficient neural architecture search via parameter sharing," *arXiv preprint arXiv:1802.03268*, 2018.

[12] E. Real, A. Aggarwal, Y. Huang, and Q. V. Le, "Regularized evolution for image classifier architecture search," in *AAAI*, vol. 33, 2019.

[13] C. Liu, L.-C. Chen, F. Schroff, H. Adam, W. Hua, A. L. Yuille, and L. Fei-Fei, "Auto-deeplab: Hierarchical neural architecture search for semantic image segmentation," in *IEEE CVPR*, 2019.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *NeurIPS*, 2017.

[15] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International Conference on Machine Learning*. PMLR, 2019.

[16] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring face images with exemplars," 2014.

[17] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *ECCV*, 2016.

[18] G. Boracchi and A. Foi, "Modeling the performance of image restoration from motion blur," *IEEE TIP*, vol. 21, no. 8, 2012.

[19] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *IEEE CVPR*, 2018.

[20] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *IEEE CVPR*, 2018.

[21] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *IEEE CVPR*, 2019.

[22] K. Wei, Y. Fu, J. Yang, and H. Huang, "A physics-based noise formation model for extreme low-light raw denoising," in *IEEE CVPR*, 2020.

[23] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *IEEE ICCV*, 2015.

[24] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE, 2018.

[25] K. Xu, X. Yang, B. Yin, and R. W. Lau, "Learning to restore low-light images via decomposition-and-enhancement," in *IEEE CVPR*, 2020.

[26] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *IEEE CVPR*, 2019.