Learning-based Methods for Single Image Restoration and Translation

He Zhang

Adobe

September 17, 2019

Outline

Self Introduction and Research Overview

Self Introduction

He Zhang, Research Scientist at Adobe

Research:

1. Image Enhancement
2. Image Compositing
3. Sparse and Low-rank Representation

Specialty Skills

I was a professional athlete for 100m and 200m.

I was a second-class national athlete in China.

Presentation Overview

Motivations

Input De-rained results

Presentation Overview

Single Image De-raining

Remove rain-streaks from a single image.

Input De-rained results

Presentation Overview (1)

Single Image Dehazing

Remove haze from a single image.

Before Dehazing

After Dehazing

Presentation Overview (2)

Thermal-to-visible Face Synthesis

Translate the thermal image into the visible domain.

Polar S_0 S_1 S_2 Proposed Target

Single Image De-raining

Problem Formulation

$$\mathbf{y} = \mathbf{y}_c + \mathbf{y}_r, \quad (1)$$

$\mathbf{y}$: Rainy image

$\mathbf{y}_c$: Target image (Clean image)

$\mathbf{y}_r$: Rain-streak components

(a) (b) (c)

Rain streaks removal from a single image. A rainy image (a) can be viewed as the superposition of a clean background image (b) and a rain streak image (c).

Related Works

Prior-based

Develop de-raining methods based on **different priors**. (e.g. sparsity-prior.)

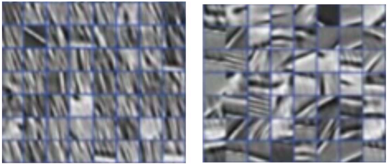
Deep Learning based

CNN de-raining methods via leveraging synthetic datasets to learn the mapping: **Rainy image** $\rightarrow$ **Clean image**.

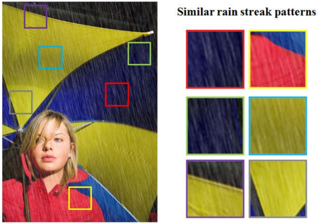
Prior-based Methods

Sparsity prior [TIP'12][ICCV'15][ICCV'17]: Learn two different dictionaries to sparsely represent clean image and rain-streak components separately.

Low-rank prior [ICCV'13][WACV'17]: Leverage patch-rank as a prior to characterize unpredictable rain-streak patterns.



Sparsity prior:
(a) Rain Dict; (b) Non-rain Dict



Low-rank Prior

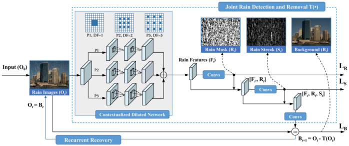
Deep Learning based Methods

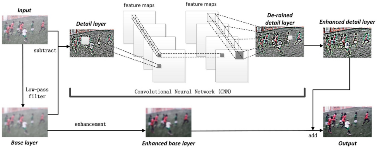
CNN [TIP’17]: Directly learn the mapping between rainy and clean image via detail layers.

DDN [CVPR’17]: Deep detail network to directly reduce the mapping range from input to output (operate on the high-frequency domain).

JORDER: [CVPR’17]: Deep learning method for joint rain detection and removal.

Many new methods now !!!.





Our Contributions

* A **density-aware multi-stream** [a] framework is proposed to remove rain-streaks with different scales, shapes and densities.

[a] Accpted in CVPR'18

Observation

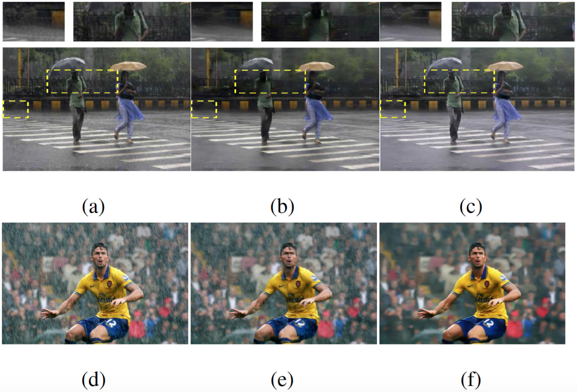


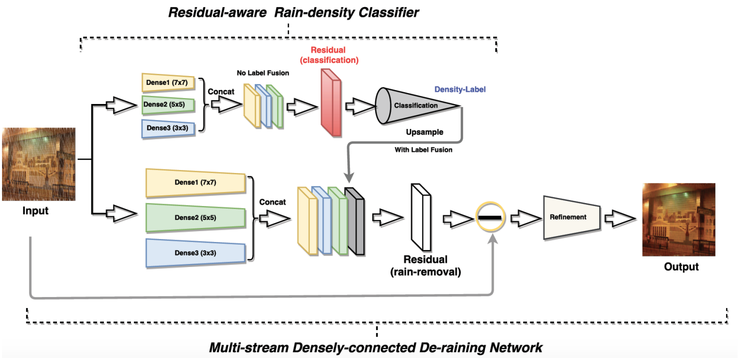
Image de-raining results. (a) Input rainy image. (b) Result from Fu *et al.* (c) DID-MDN. (d) Input rainy image. (e) Result from Li *et al.* (f) DID-MDN.

Observation (1)

(a) (b)

Sample images containing rain-streaks with various scales and shapes.(a) contains smaller rain-streaks, (b) contains longer rain-streaks.

Proposed Method



The proposed network contains two modules:
(a) residual-aware rain-density classifier.
(b) multi-stream densely-connected de-raining network.

This is optimized via Euclidean loss.

Training Details

Datasets

Synthesized {Rainy/Clean} based on image-degradation models via different rain-mask created by Photoshop.

TrainA: Synthesize using natural images with 3 density-label (heavy, medium and light) and in total 12000 samples (each with 4000).

TestA: Synthesize using natural images with 3 density-label (heavy, medium and light) and in total 1200 samples (each with 400).

TestB: Synthesized 1000 samples from CVPR'17 paper.

Synthetic Images

Quantitative results evaluated in terms of average SSIM and PSNR (dB).

	Input	DSC (ICCV'15)	GMM (CVPR'16)	CNN (TIP'17)	JORDER (CVPR'17)	DDN (CVPR'17)	JBO (ICCV'17)	DID-MDN
Test1	0.7781/21.15	0.7896/21.44	0.8352/22.75	0.8422/22.07	0.8622/24.32	0.8978/ 27.33	0.8522/23.05	**0.9087**/ **27.95**
Test2	0.7695/19.31	0.7825/20.08	0.8105/20.66	0.8289/19.73	0.8405/22.26	0.8851/25.63	0.8356/22.45	**0.9092**/ **26.0745**

Real Images

Input *JORDER (CVPR'17)* *DDN (CVPR'17)* *JBO (ICCV'17)* *DID-MDN*

Rain-streak removal results on sample real-world images.

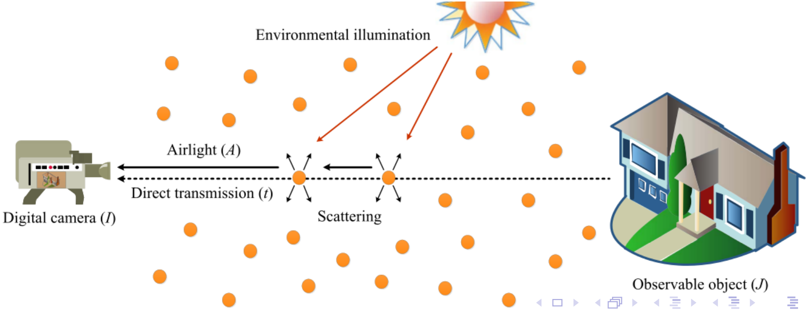
Single Image Dehazing

Problem Formulation

The observation model is:

$$I = \overbrace{J * t}^{\textbf{Direct Attuation}} + A\underbrace{(1 - t)}_{\textbf{Air-light}},$$

I: Hazy image | J: Target image
A: Atmospheric light | t: Transmission map
($t = e^{-\beta d}$, β: attenuation coefficient; d is the depth.)



Goal

The observation model is:

$$I = J * t + A(1 - t), \tag{2}$$

I: Hazy image
J: Target image
A: Atmospheric light
t: Transmission map

Given I, estimate J

$$\hat{J} = \frac{I - \hat{A}(1 - \hat{t})}{\hat{t}}$$

Alternative Goal: Estimate $\hat{A}$ and $\hat{t}$

Related Work

Common Approach

Accurate transmission map $\rightarrow$ Better dehazing

(Concentrate on estimating the transmission map t; Empirical rule to estimate atmospheric light A.)

These methods (estimating transmission map) can be divided into two separate groups: **Prior-based** and **Learning-based**.

Prior-based

Develop estimation methods based on empirical observation. (e.g. hazy image loose color contrast.)

Learning-based

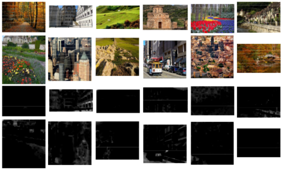
CNN estimation methods via leveraging synthetic datasets. Hazy image $\rightarrow$ Transmission map.

Prior-based Methods

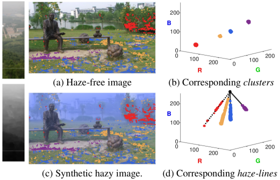
Dark-channel prior [CVPR'09]: Outdoor objects in haze-free weather have at least one color channel that is significantly dark,

Color-line prior [TOG'14]: Small image patches typically exhibit a one-dimensional distribution in the RGB color space.

Haze-line prior [CVPR'16]: Colors of a haze-free image can be well approximated by a few hundred distinct colors.



(CVPR'09)



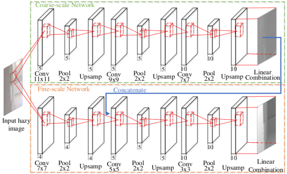
(CVPR'16)

Learning-based Methods

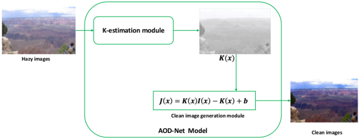
Multi-scale net [ECCV'16]: A coarse to fine multi-scale structure to estimate the transmission map.

AOD-net [ICCV'17]: Directly generates the clean image through a light-weight CNN via using linear transformation to embed t and A as one variable.

Many new methods now !!!.



ECCV'16



(a) The diagram of AOD-Net

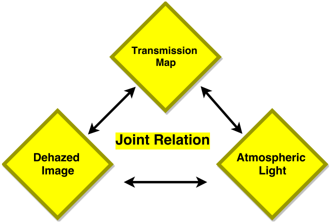
ICCV'17

Challenges for Leaning-based Methods

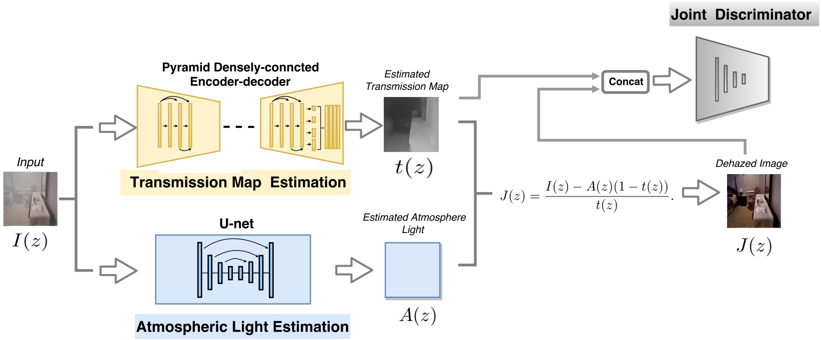
Challenges

(1) *Inaccuracies in the estimation of transmission map t*
$\rightarrow$ Low quality de-hazed result.

(2) *Non end-to-end learning*
$\rightarrow$ Unable to capture inherent relations among transmission map t, atmospheric light A and dehazed image J.



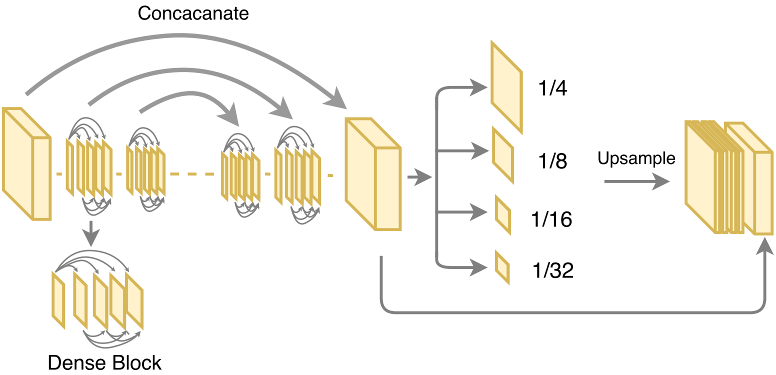
Densely-connected Pyramid Dehazing Network



Our contributions: *Accepted in CVPR'18*

1. End-to-end learning via embedding dehazing model into the network.
2. Pyramid densely connected encoder-decoder for estimating the transmission map.
3. A novel edge-preserving loss to avoid halo-artifacts.
4. Joint discrimanter to decide whether paired samples (t and J) are real or fake.

Pyramid Densely-connected Transmission Estimation Network



An overview of the proposed pyramid densely connected transmission map estimation network.

Why Dense Block as Basic Structure?

Advantages

(1) Features from different levels are used for prediction.

(2) Improved flow of information and gradients.

(3) Less parameters (features re-use).

SSIM:0.8223 *SSIM:0.8882* *SSIM:1*

U-net DED Target

Why Multi-level Pooling?

Unsolved issue

'Global' structural information of objects with different scales is **lost**.

Solution

Embed information from **different scales** to make the estimation decision.

SSIM:0.8882 *SSIM:0.9119* *SSIM:1*

DED DED-MLP Target

Halo artifacts

Halo artifacts.

L2 loss tends to blur the output (transmission map).
$\rightarrow$ halo artifacts in the dehazed image.

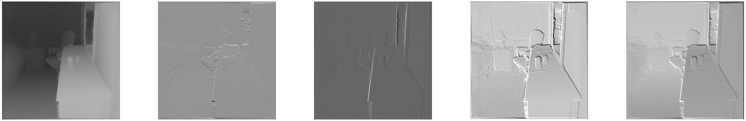
Solution for halo artifacts

Edge information should be considered in the loss function.

Edge-Preserving Loss

Observations

* Edges correspond to the discontinuities in the image intensities.
* Shallow layers of CNN can function as edge detector.



(a) (b) (c) (d) (e)

Feature visualization for gradient operator and low-level features. (a) Input transmission map. (b) Horizontal gradient output. (c) Vertical gradient output. (d) and (e) are visualization of two feature maps from relu1_2 of VGG-16.

Edge-Preserving Loss (2)

EP-Loss

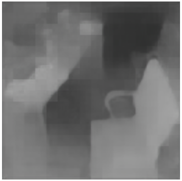
$$L^E = \lambda_{E,l_2} L_{E,l_2} + \lambda_{E,g} L_{E,g} + \lambda_{E,f} L_{E,f}, \tag{3}$$

L_{E,l_2}: L2 loss, $L_{E,f}$: CNN feature loss.
$L_{E,g}$: two-directional (horizontal and vertical) gradient loss.

SSIM:0.9119

DED-MLP

SSIM:0.9201



DED-MLP-GRA

SSIM:0.9213

DED-MLP-PER

SSIM:1

Target

Joint Discriminator

Motivation

Structure of tiny objects and objects with larger depth are still missing.

Solution

Leverage the strong capabilities of generative adversarial network to synthesize the missing structure details.

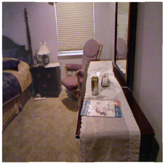
SSIM:0.9213 *SSIM:0.9283* *SSIM:1*

DED-MLP-PER DCPDN Target

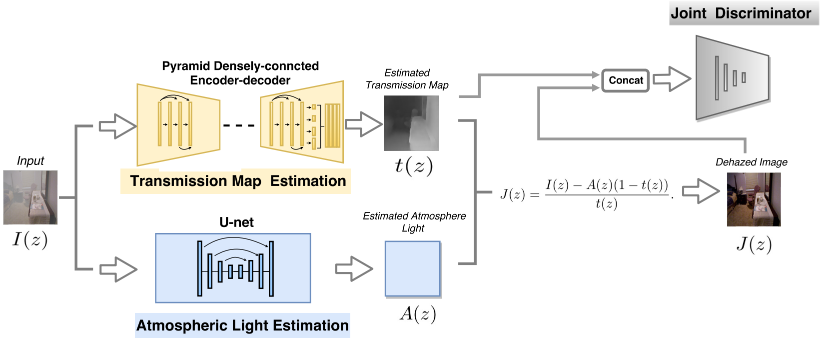
Joint Discriminator (2)

Structural information between $\hat{t}$ and $\hat{J}$ are highly correlated.
A joint discriminator to learn a joint distribution to decide whether the corresponding pairs (*transmission map*, *dehazed image*) are real or fake.

$$\begin{aligned} \min_{G_t, G_d} \max_{D_{joint}} \quad & \mathbb{E}_{I \sim p_{data(I)}} [\log(1 - D_{joint}(G_t(I)))] + \\ & \mathbb{E}_{I \sim p_{data(I)}} [\log(1 - D_{joint}(G_d(I)))] + \\ & \mathbb{E}_{t, J \sim p_{data(t,J)}} [\log D_{joint}(t, J))]. \end{aligned} \tag{4}$$




Overall Loss

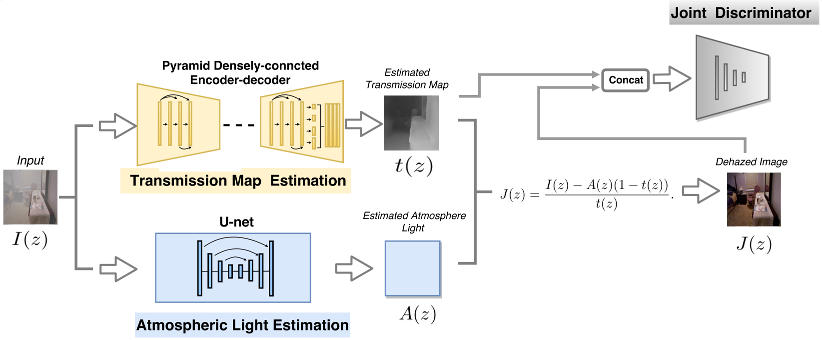


Loss function

$$L = L^t + L^a + L^d + \lambda_j L^j, \tag{5}$$

L^t: Edge-preserving loss L^E; L^a: $L2$ loss in predicting the atmospheric light; L^d: $L2$ loss represents the dehazing loss; L^j: Joint discriminator loss

Stage-wise Training



Solution

Initialize different parts of network to '**better**' conditions and then optimize all parts together in the end.
(*Learn each module progressively and then optimize all in the end.*)

Training Details

Datasets

Synthesized {Hazy /Clean /Transmission Map /Atmosphere Light} on Image-degradation models via changing A and β using existing depth datasets.

TrainA: Synthesize using NYU-depth2 datasets with 4000 samples.

TestA: Synthesize using NYU-depth2 datasets with 400 samples.

TestB: Synthesize using Middlebury stereo database (40) and also the Sun3D dataset (160) with 200 samples.

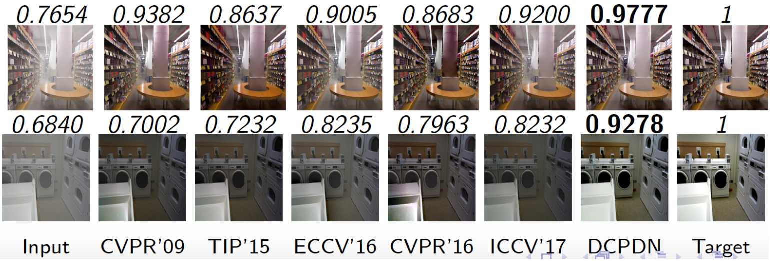
Synthetic Images

Quantitative SSIM results on the synthetic **TestA** dataset.

	Input	He. *et al.* (CVPR'09)	Zhu. *et al.* (TIP'15)	Ren. *et al.* (ECCV'16)	Berman. *et al.* (CVPR'16)	Li. *et al.* (ICCV'17)	DCPDN
Transmission	N/A	0.8739	0.8326	N/A	0.8675	N/A	**0.9776**
Image	0.7041	0.8642	0.8567	0.8203	0.7959	0.8842	**0.9560**

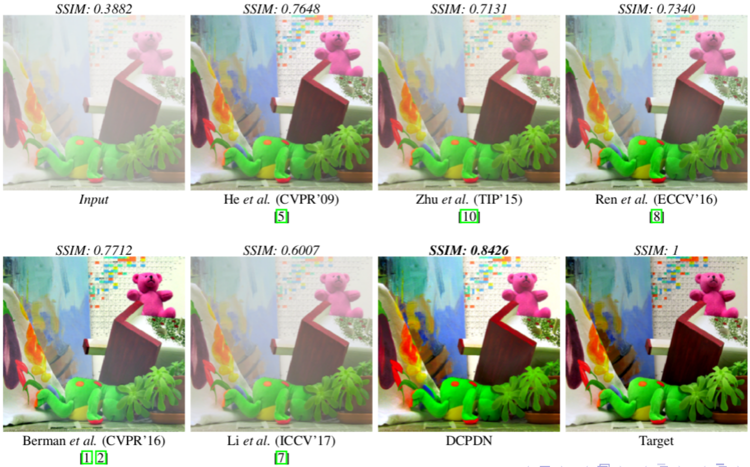
Quantitative SSIM results on the synthetic **TestB** dataset.

	Input	He. *et al.* (CVPR'09)	Zhu. *et al.* (TIP'15)	Ren. *et al.* (ECCV'16)	Berman. *et al.* (CVPR'16)	Li. *et al.* (ICCV'17)	DCPDN
Transmission	N/A	0.8593	0.8454	N/A	0.8769	N/A	**0.9352**
Image	0.6593	0.7890	0.8253	0.7724	0.7597	0.8325	**0.8746**



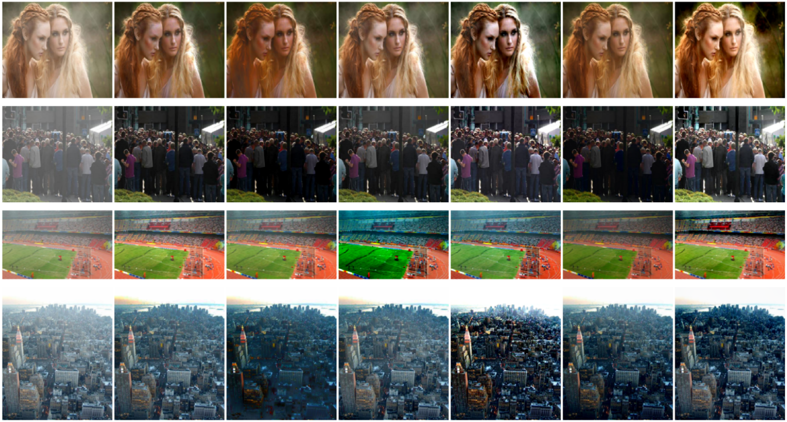
Synthetic Images (2)

Dehazed visual comparisons for results of synthetic image used by previous methods.



Real Images

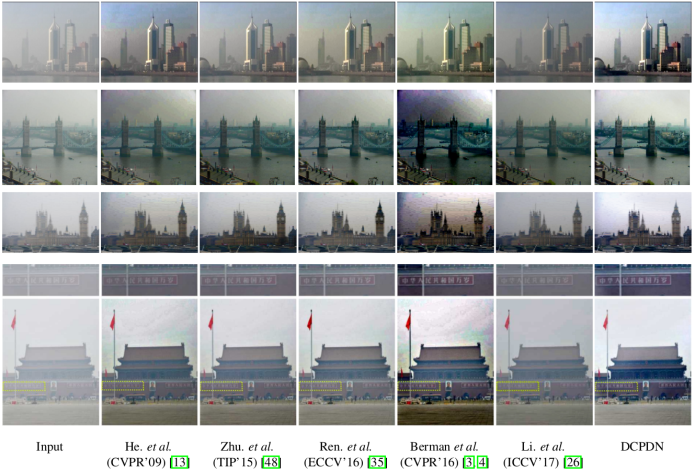
Dehazed visual comparisons for results of real images used by previous methods.



Input | He. *et al.* (CVPR'09) [13] | Zhu. *et al.* (TIP'15) [48] | Ren. *et al.* (ECCV'16)[35] | Berman. *et al.* (CVPR'16) [3] [4] | Li. *et al.* (ICCV'17) [26] | DCPDN

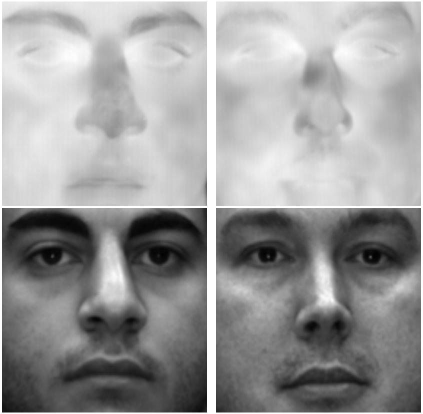
Real Images (2)

Dehazed visual comparisons for results of real images from the Internet.



Thermal-Visible Face Synthesis and Verification

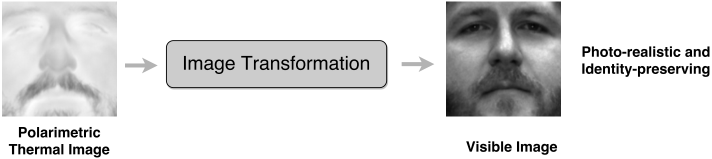
Motivations



Large **domain discrepancy** makes cross-domain face recognition quite a challenging problem for human-examiners and computer vision algorithms.

Goal

Learn a transformation to transfer the **polarimetric thermal image** to the **visible domain** and make sure the generated images are photo-realistic and identity-preserving.



What is Polarimetric Thermal Image?

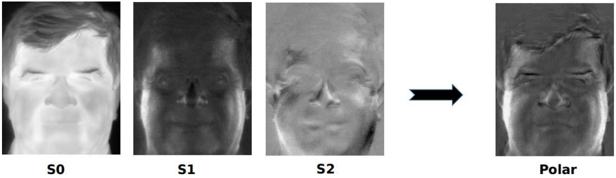
Composed of three channels: S0, S1 and S2:

S0 $\rightarrow$ conventional thermal image.

S1 $\rightarrow$ the horizontal polarization-state image.

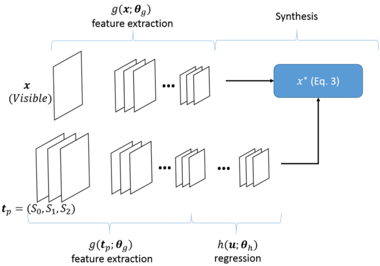
S2 $\rightarrow$ the vertical polarization-state image.

S1 and S2 complements S0 by providing additional textural and geometric details.



Related Works

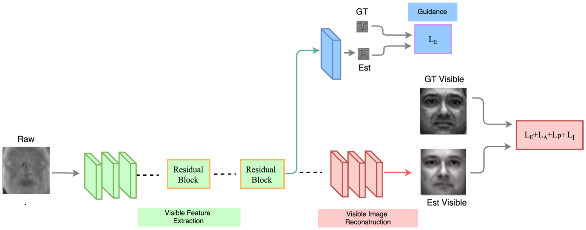
* Make sure the visible image and polarimetric thermal image contain the **same feature representations**.
* Project the estimated features back to the image domain.



Riggan, Benjamin et al. "Estimation of visible spectrum faces from polarimetric thermal faces." IEEE BTAS,2016

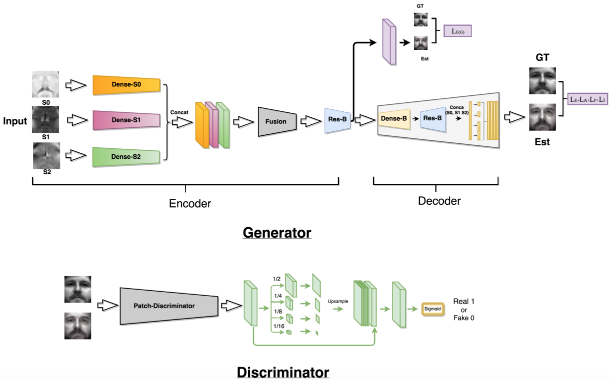
Related Works (1)

* End-to-end learning via input-level fusion.



Zhang, He, et al. "Generative adversarial network-based synthesis of visible faces from polarimetrie thermal faces." Biometrics (IJCB), 2017 IEEE International Joint Conference on. IEEE, 2017.

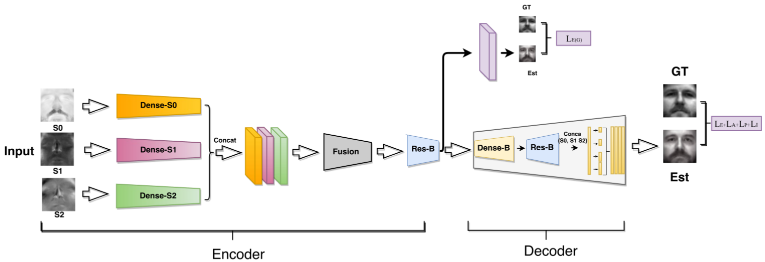
Proposed Method



Our contributions: *Accepted in IJCV*

1. Face synthesis with GAN (multi-stream generator and multi-scale discriminator).
2. An extended dataset consisting of 111 subjects is collected.

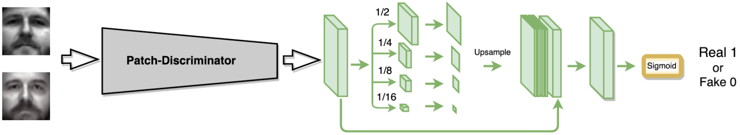
Feature-level Fusion



Generator

Each encoder inherently learns to characterize different geometric and texture information that is captured in the Stokes images.

Multi-scale Discriminator



Discriminator

Leverage information from different scales to make the decision.

Training Details

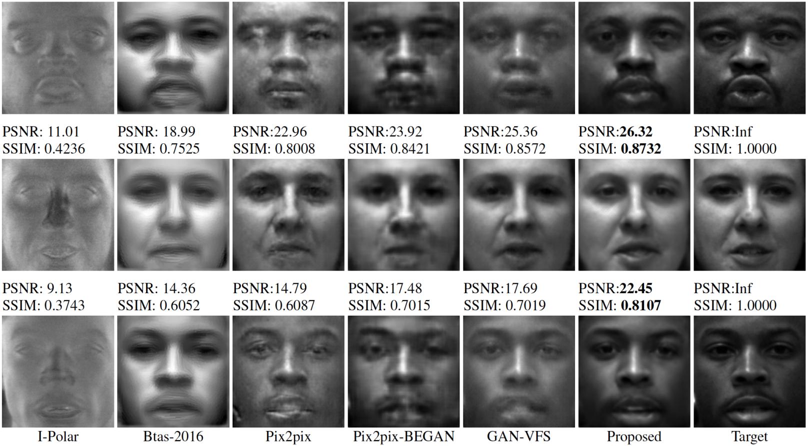
Datasets Discription

(a) 111 subjects.
(b) 4 modalities (S0, S1, S2, Visible).
(c) 12 images per subject with baseline expression and 18 images per subject with various expressions.

Train and Test Samples

Train: 680 images from randomly selected 85 subjects.
Test: 208 images from the other 26 subjects.

Experimental Results



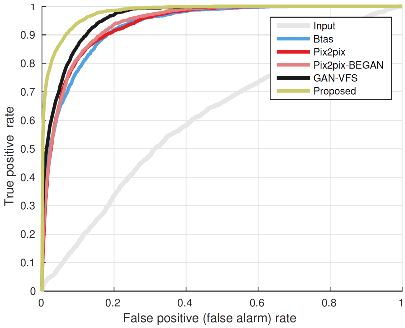
Visual comparisons compared with state-of-the-art methods.

Experimental Results (1)

Quantitative Performance: PSNR and SSIM

	I-Polar	Btas-2016 [46]	Pix2pix [19]	Pix2pix-BEGAN [19, 1]	GAN-VFS [67]	Proposed
PSNR (dB)	10.88	15.82	17.82	18.28	18.58	**19.18**
SSIM	0.4467	0.6854	0.6828	0.7214	0.7283	**0.7340**

ROC Curve



Related Research Topic (More)

Unconstrained Face Detection Dataset (UFDD) in Severe Weather Conditions

* A new dataset of face images that involve weather-based degradations, motion blur, focus blur and several others.
* A considerable gap in the performance of state-of-the-art detectors and real-world requirements.

(Pushing the limits of unconstrained face detection: a challenge dataset and baseline results, BTAS-2018)

Single Image Face De-blurring

* Learn weights differently for different classes/parts in the face.

(Deblurring Face Images using Uncertainty Guided Multi-Stream Semantic Networks, arxiv)

Conclusion

Conclusion

Single Image De-raining

*Density-aware de-raining method is proposed.

Single Image Dehazing

*End-to-end dehazing method is proposed.

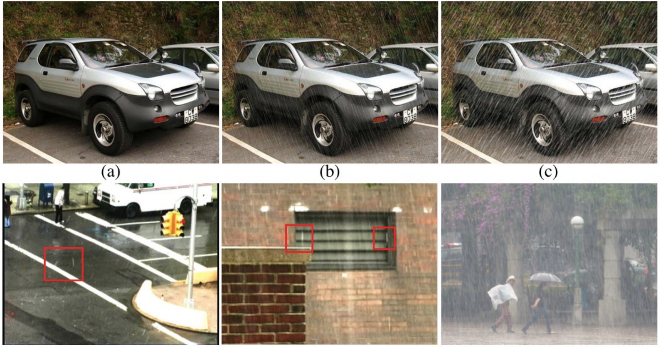
Thermal-to-Visible Face Synthesis

*Feature-level fusion network is proposed.

Future Directions

Improve the Quality of Synthetic Images

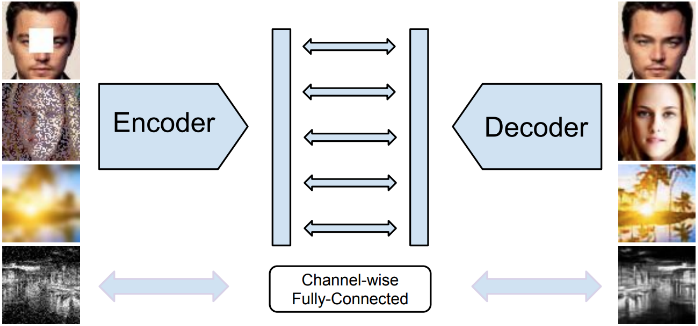
How to improve quality of synthetic images to make it realistic for single image de-raining and single image dehazing problem.



(a) (b) (c)
(d) (e) (f)

Design All-in-one Image Restoration Models

Explore the possibility of designing a model/framework which is able to address all image restoration problems together.



Thanks!