SIGGRAPH ASIA 2019 BRISBANE

DeepRemaster: Temporal Source-Reference Attention Networks for Comprehensive Video Enhancement

Satoshi Iizuka Edgar Simo-Serra

sa2019.siggraph.org

Background

- Vintage film is deteriorated
  - Noise, blur, and low contrast
  - Black and white or low quality colors

"Oliver Twist" (1933)

"A-Bomb Blast Effects" (1952)

- Digital remastering is a challenging task
  - Conducted manually by experts
  - Requires a significant amount of both time and money

Seven Samurai (1954)

Our Goal

- Semi-automatically remastering of vintage films
  - This includes restoration, enhancement, and colorization

Related Work

- Image/video restoration
  - Gaussian noise [Dabov+ ’07, Maggioni+ ’12 ’14, Lefkimmiatis ’18]
  - JPEG noise [Zhang+ ’17]
  - Blur [Shi+ ’16]

Gaussian Noise

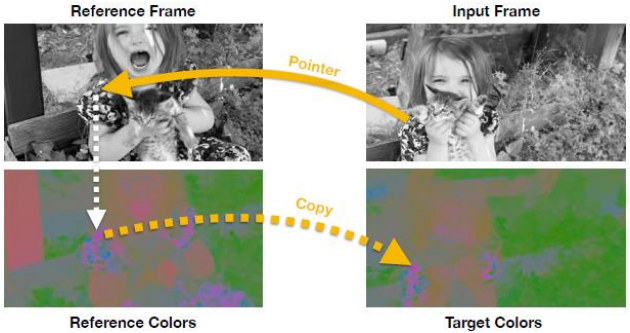
- Image Colorization
  - Scribble-based [Levin+ 2004; Yatziv+ ’04; An+ ’09; Xu+ ’13; Endo+ ’16; Zhang+ ’17]
  - Reference-based [Chia+ ’11; Gupta+ ’12; He+ ’18]
  - Automatic [Iizuka+ ’16; Larsson+ ’16; Zhang+ ’16]

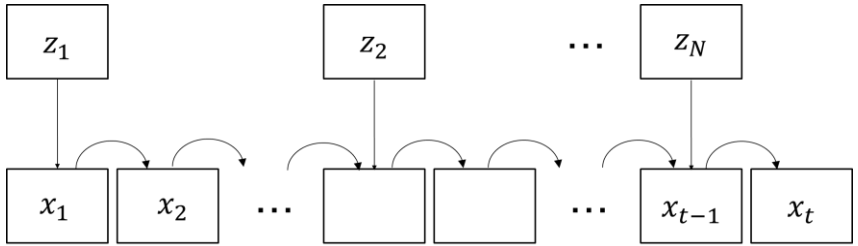
[Levin+ ’04]

[Zhang+ ’17]

Related Work

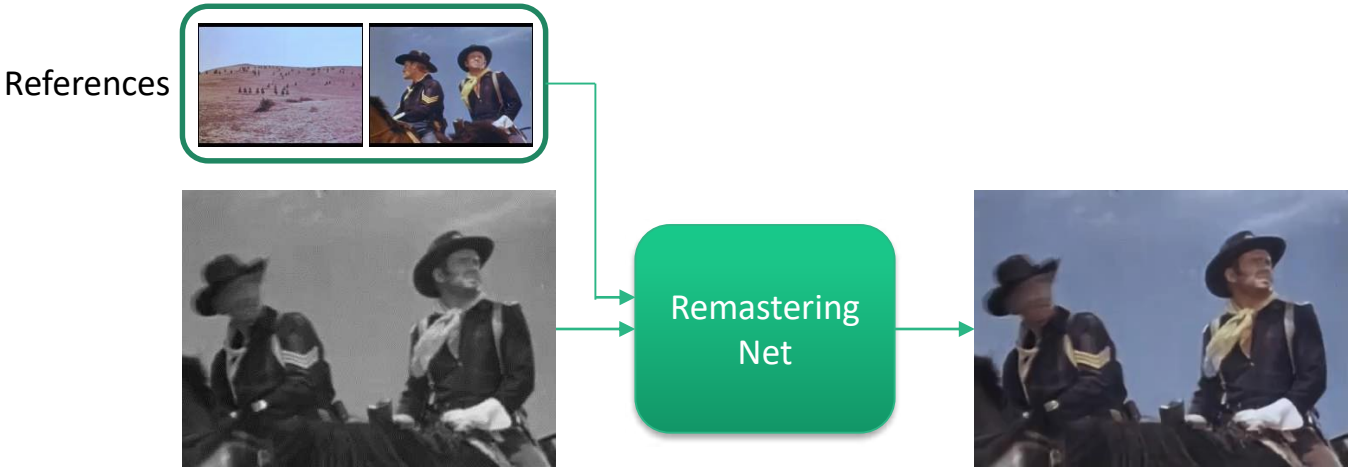
- Reference-based video colorization
  - Recurrent neural networks [Liu+ ’18; Vondrick+ ’18]
  - Processes a video by propagating color frame-by-frame
  - Cannot propagate between scene changes
  - Continues amplifying errors



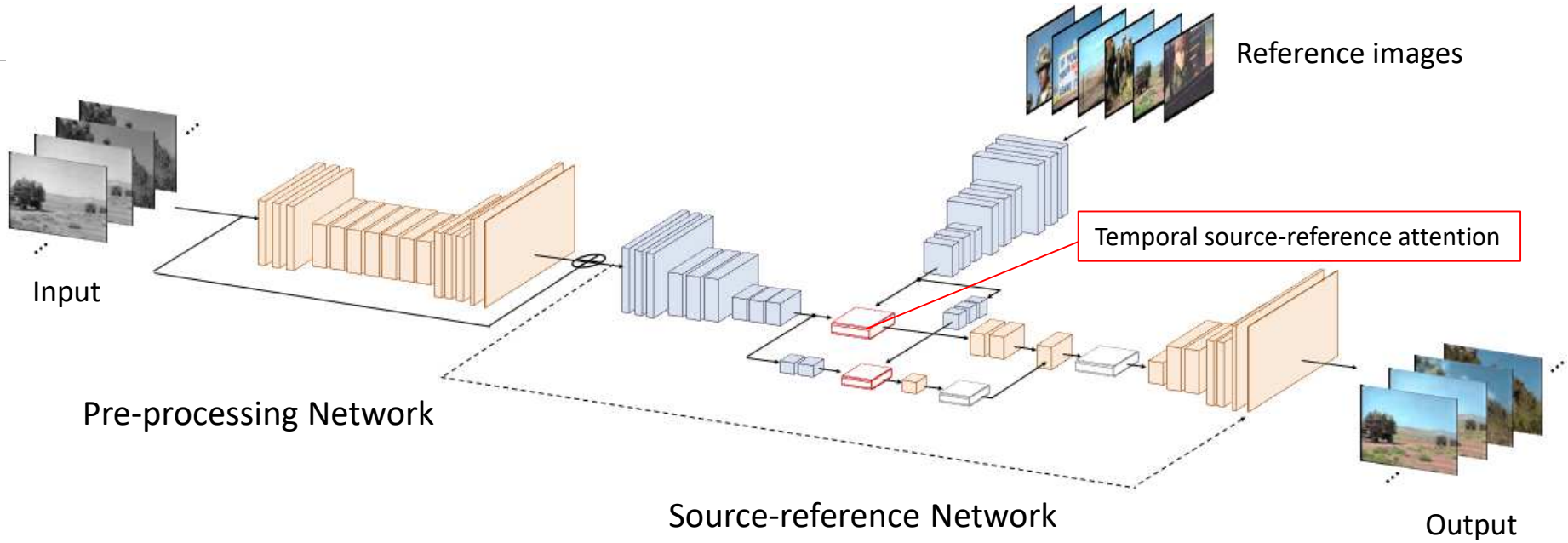


Our Method

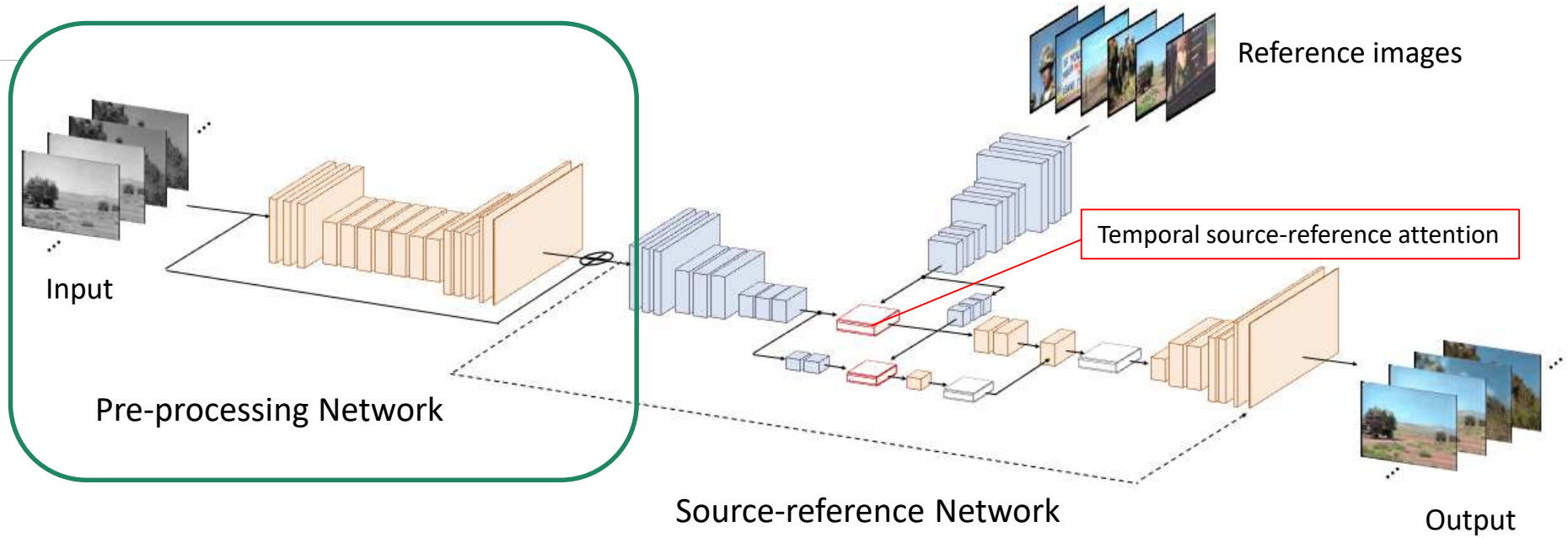
- Model based on spatial and temporal convolutions
  - Automatic noise removal, super-resolution, and contrast adjustment
- Semi-automatic colorization source-reference attention
  - Can colorize a video using an arbitrary number of reference images



Our Network

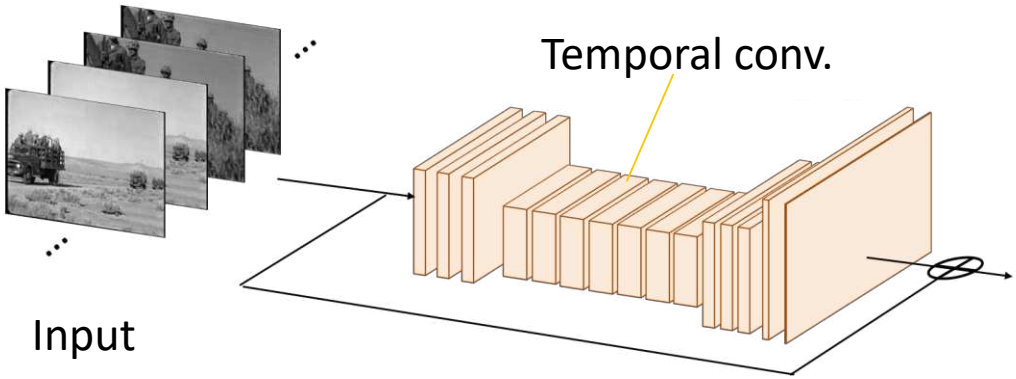

Reference images
Temporal source-reference attention
Input
Pre-processing Network
Source-reference Network
Output

Our Network

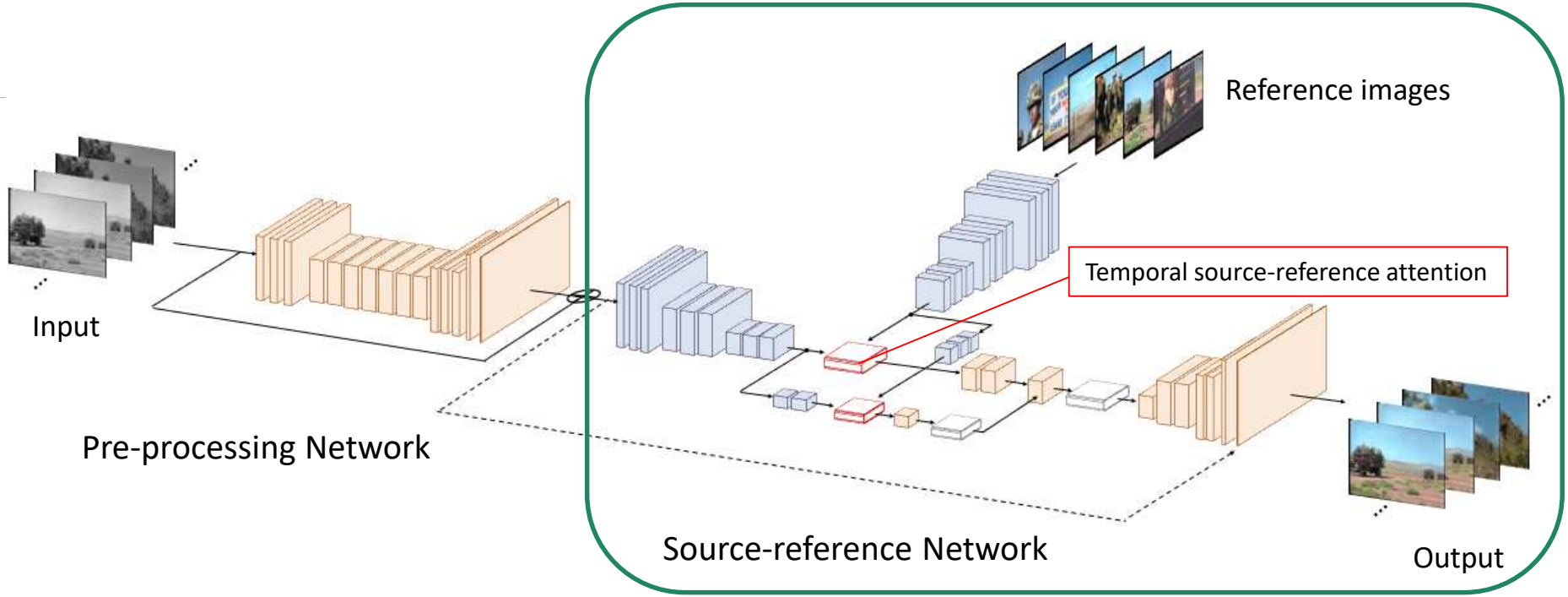

Reference images
Temporal source-reference attention
Input
Pre-processing Network
Source-reference Network
Output

Pre-processing Network

- Removes artifacts and noise from the input grayscale video
- Formed exclusively by temporal convolutions



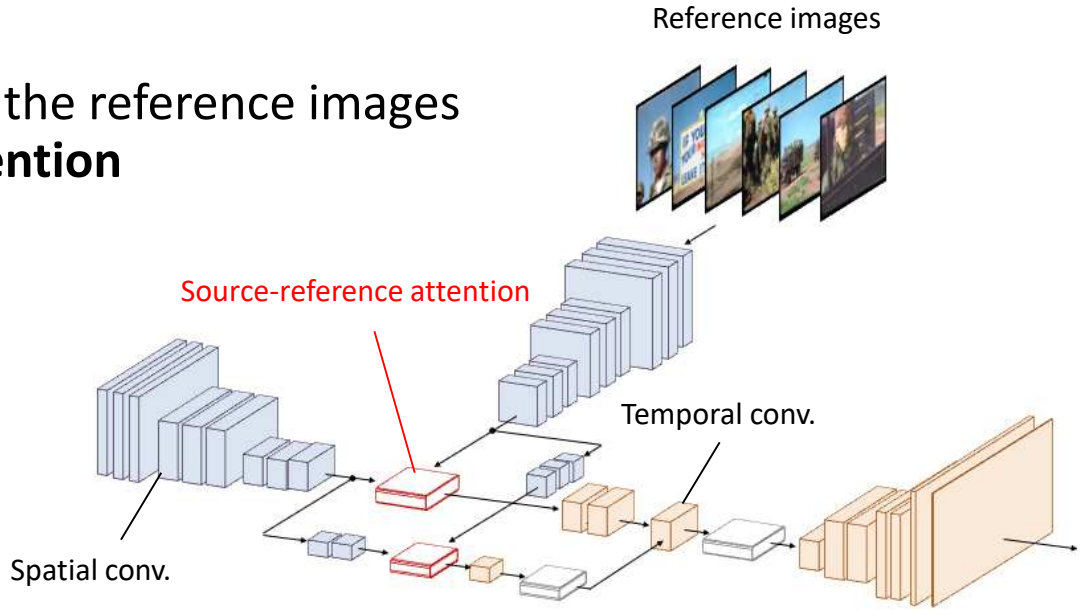
Our Network


Reference images
Temporal source-reference attention
Input
Pre-processing Network
Source-reference Network
Output

Source-reference Network

- Takes the output of the pre-processing network along with an arbitrary number of reference color images

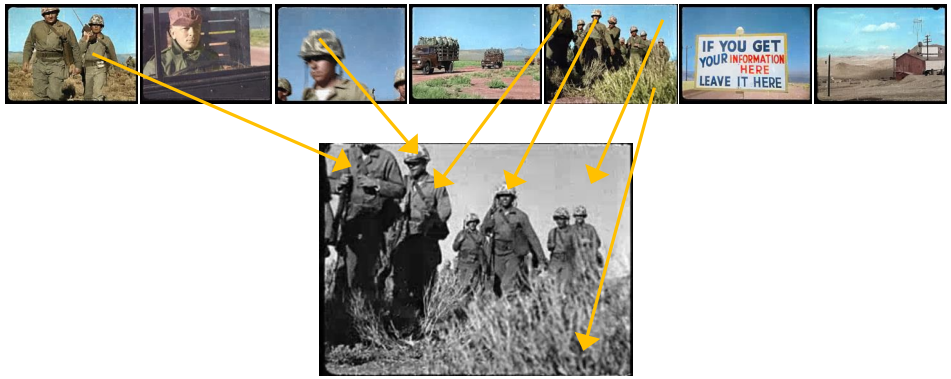
- Colorizes the frames based on the reference images by using **source-reference attention**



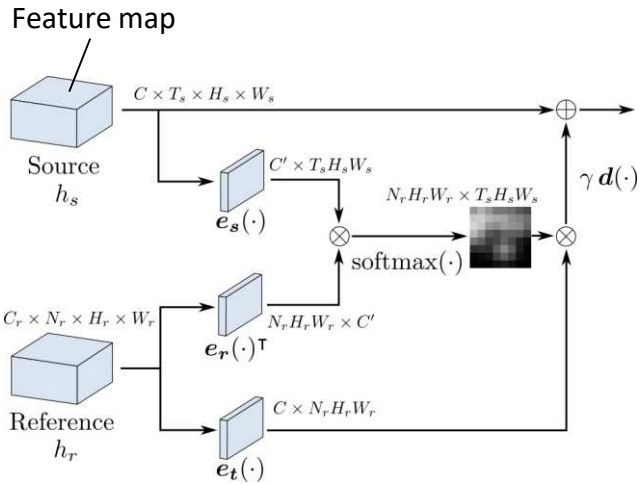
Temporal Source-reference Attention

- Compute similarity between the source images and reference images
  - Actually computed on feature maps

References

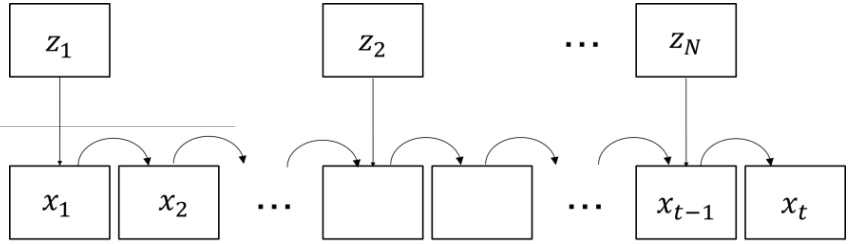


Input frame



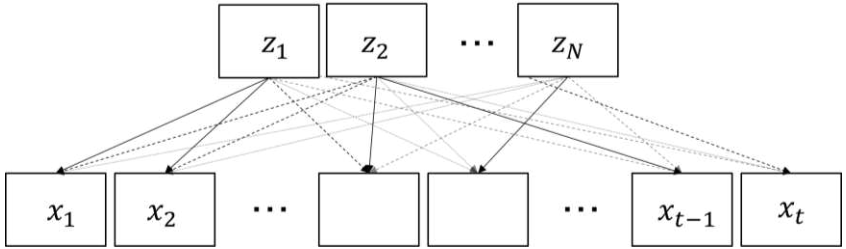
source-reference attention

Advantages of Source-Reference Attention



Recursion-based network

- Cannot form long-term dependencies
- Temporal consistency is lost when a new reference is used
- Require precise scene segmentation



Our temporal source-reference attention

Can use all the color reference information for colorization

Optimization

- Combination of two L1 losses
  - Fully supervised learning
  - Uses ADADELTA[Zeiler ’12] for optimization

- Objective function:

$$\arg\min_{\theta,\phi} \mathbb{E}_{(x,y_l,y_{ab},z)\in\mathcal{D}} \|P(x;\theta) - y_l\| + \beta \|S(P(x;\theta), z;\phi) - y_{ab}\|$$

Output of pre-processing network

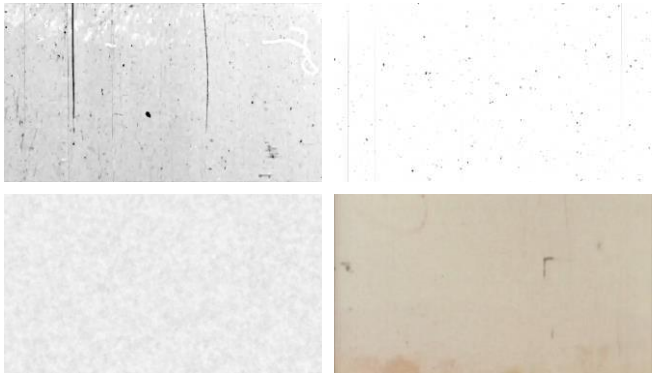
Ground truth luminance

Output of source-reference network

Ground truth chrominance

Training Data Generation

- Example-based and algorithm-based deterioration
  - Example-based: scratch noise, fractal noise, dust noise, …
  - Algorithm-based: Gaussian noise, blur, low contrast
- 1200 videos from Youtube8M[Abu-El-Haji+ ’16] for training



Examples of noise data

Original

Deteriorated

Results

Comparisons

Input | [Yu+ ’18]&[Zhang+ ’17a] | [Zhang+ ’17b]&[Vondrick+ ’18] | Ours

Quantitative Result

Remastering results

Approach	Frames	# Ref.	PSNR
Zhang+[2017b]&Zhang+[2017a]	90	1	27.13
	300	5	27.31
Yu+[2018]&Zhang+[2017a]	90	1	26.43
	300	5	26.59
Zhang+[2017b]&Vondrick+[2018]	90	1	26.43
	300	5	26.60
Yu+[2018]&Vondrick+[2018]	90	1	26.85
	300	5	26.89
Ours w/o joint training	90	1	29.07
	300	5	29.23
Ours	90	1	**30.83**
	300	5	**31.14**

Quantitative Results

Restoration results

Approach	Frames	# Ref.	PSNR
[Zhang et al. 2017b]	300	-	25.08
[Yu et al. 2018]	300	-	24.49
Ours w/o skip connection	300	-	24.73
Ours	300	-	**26.13**

Colorization results

Approach	Frames	# Ref.	PSNR
[Zhang et al. 2017a]	90	1	31.28
	300	5	31.16
[Vondrick et al. 2018]	90	1	31.55
	300	5	31.70
Ours w/o temporal conv.	90	1	28.46
	300	5	28.51
Ours w/o self-attention	90	1	29.00
	300	5	28.72
Ours	90	1	**34.94**
	300	5	**36.26**

Comparisons

Input

[Yu+ ’18]&[Zhang+ ’17a]

[Zhang+ ’17b]&[Vondrick+ ’18]

Ours

Reference images
(manually created)

Results

Reference images

Attention

Input

Output

Results

Reference images

Attention

Input

Output

Results

“Isewan typhoon” (1959), the original film is provided by CBC Television Co.

Restoration Results

- Large noise removal

Input [Zhang et al. 2017b] Ours

Limitations

- Severely deteriorated film is difficult to remaster
  - Cannot fill large missing regions
- Scene with intense motion

Input

Output

Input

Output

Conclusion

- Novel single framework to tackle entire remastering task
  - Automatic noise removal, super-resolution, and contrast adjustment
  - Reference-based colorization via temporal source-reference attention
- Significant improvement with respect to existing methods
- Applicable to other reference-based image/video processing
- GitHub: https://github.com/satoshiiizuka/siggraphasia2019_remastering