Let there be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification

Satoshi Iizuka* Edgar Simo-Serra* Hiroshi Ishikawa

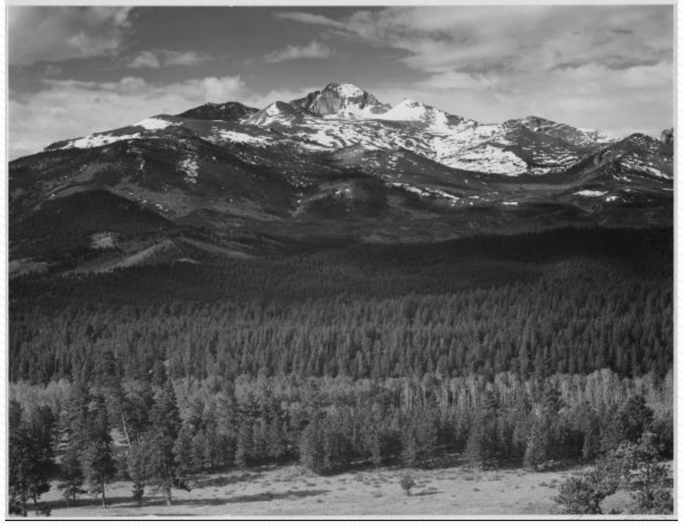
Waseda University

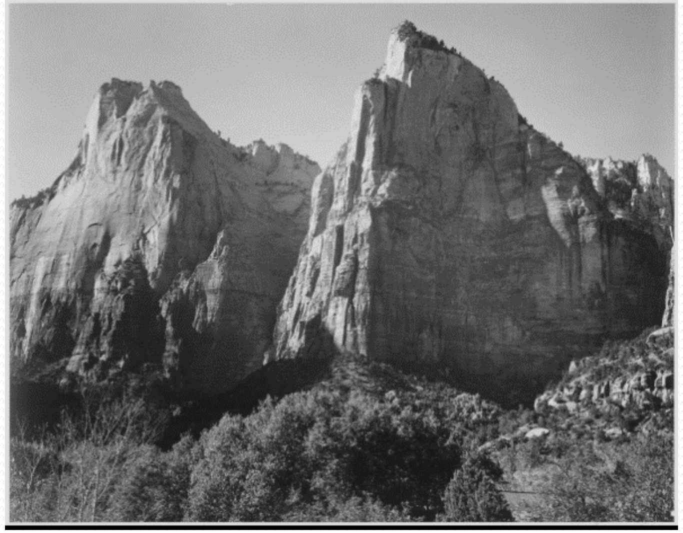
(*equal contribution)

Render the Possibilities

SIGGRAPH2016

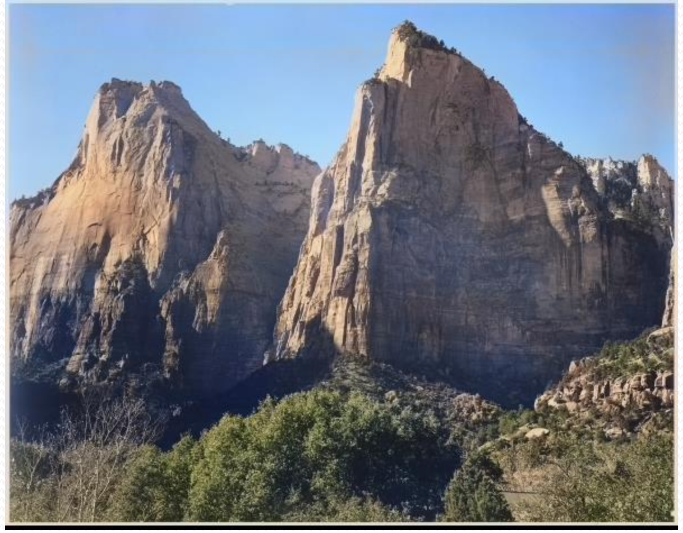
Colorization of Black-and-white Pictures





Our Goal: Fully-automatic colorization




Colorization of Old Films

Related Work

- Scribble-based [Levin+ 2004; Yatziv+ 2004; An+ 2009; Xu+ 2013; Endo+ 2016]
  - Specify colors with scribbles
  - Require manual inputs

[Levin+ 2004]

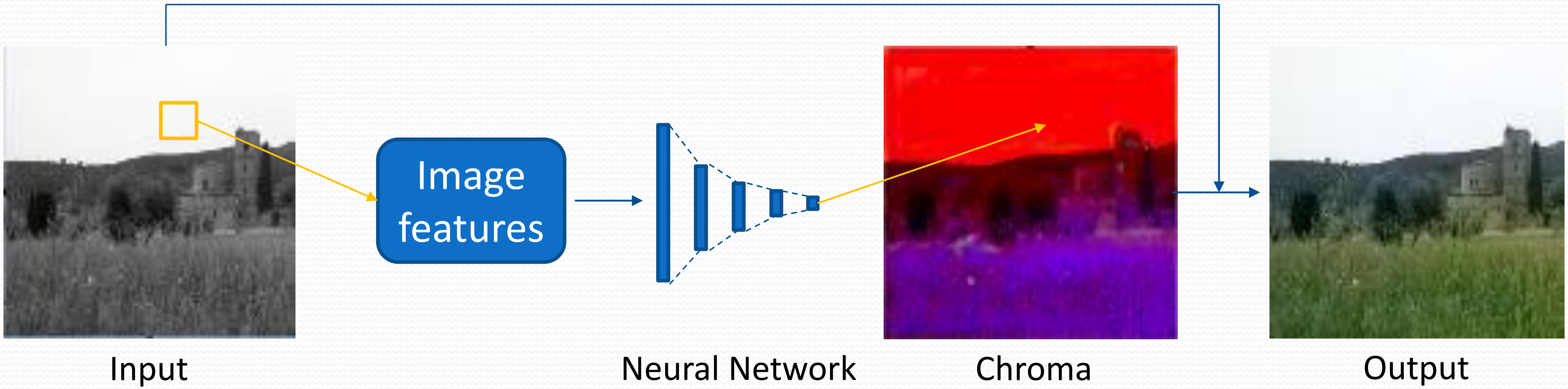
- Reference image-based [Chia+ 2011; Gupta+ 2012]
  - Transfer colors of reference images
  - Require very similar images

Input Reference Output

[Gupta+ 2012]

Related Work

- Automatic colorization with hand-crafted features [Cheng+ 2015]
  - Uses existing multiple image features
  - Computes chrominance via a shallow neural network
  - Depends on the performance of semantic segmentation
  - Only handles simple outdoor scenes



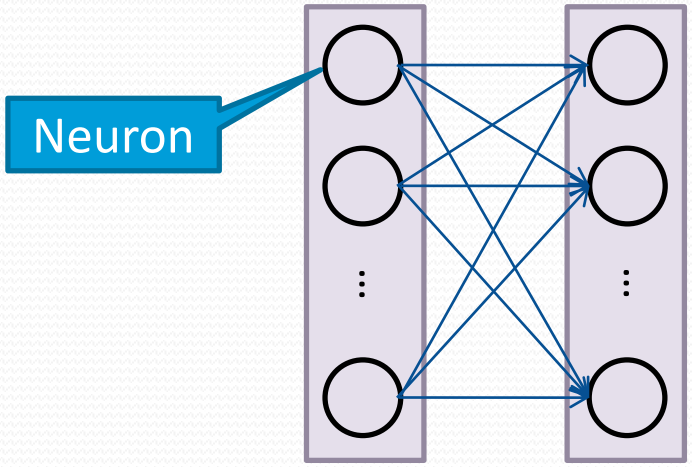
Contributions

- Novel end-to-end network that jointly learns global and local features for automatic image colorization
  - New fusion layer that elegantly merges the global and local features
  - Exploit classification labels for learning

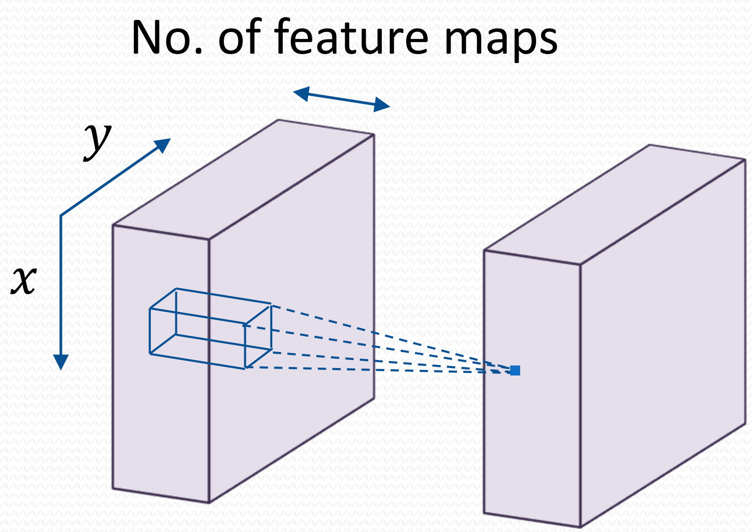



Layers of Our Model

- Fully-connected layer
  - All neurons are connected between layers
- Convolutional layer
  - Takes into account underlying spatial structure

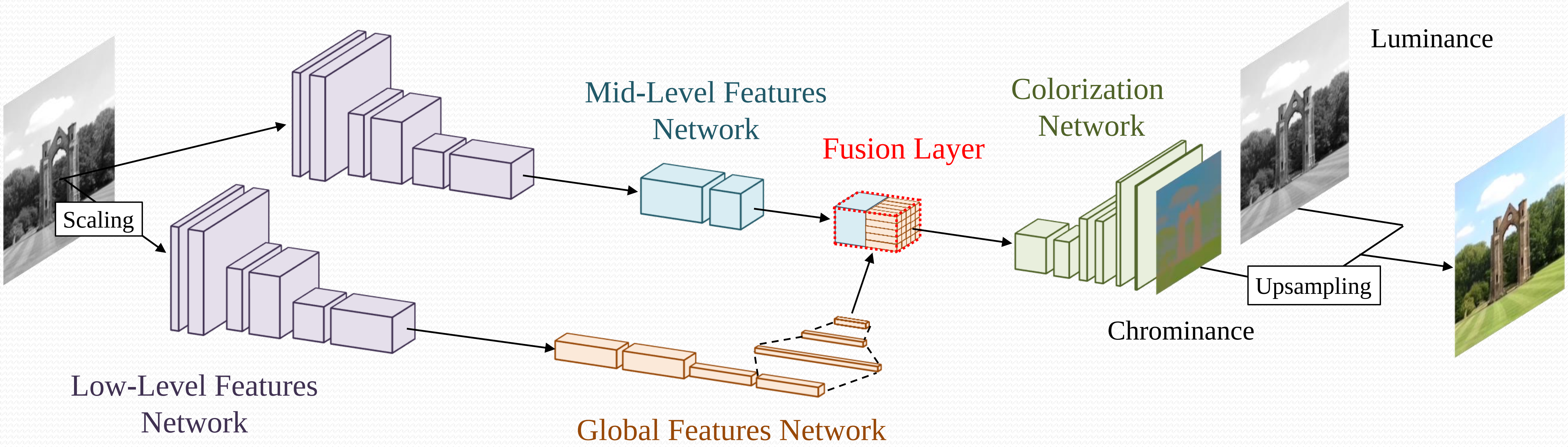


Fully-connected layer



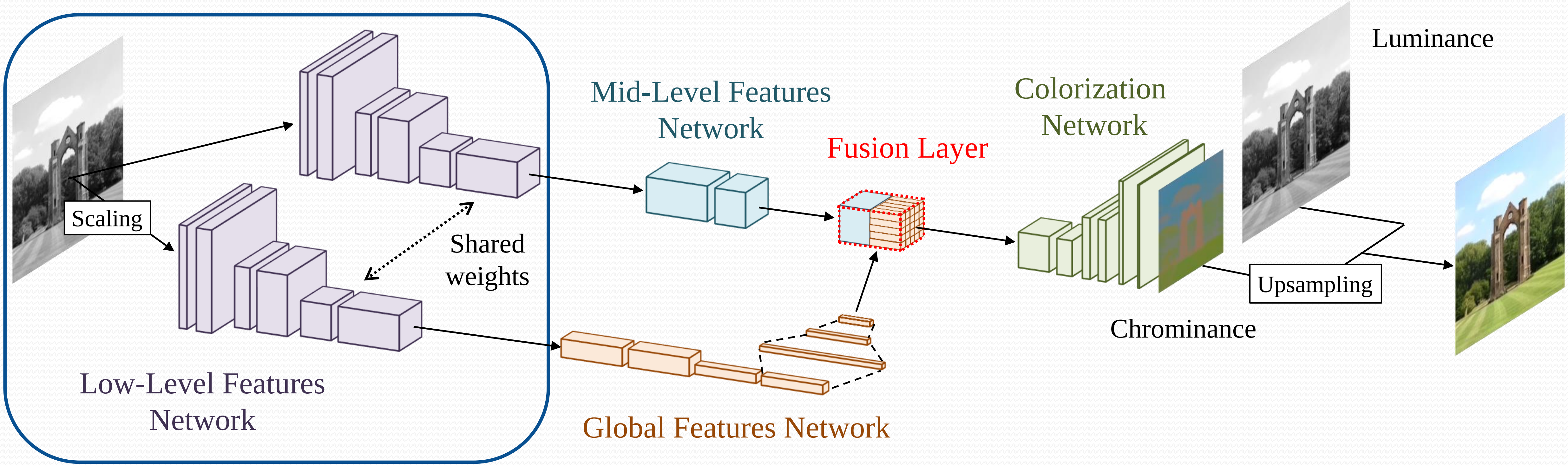
Convolutional layer

Our Model



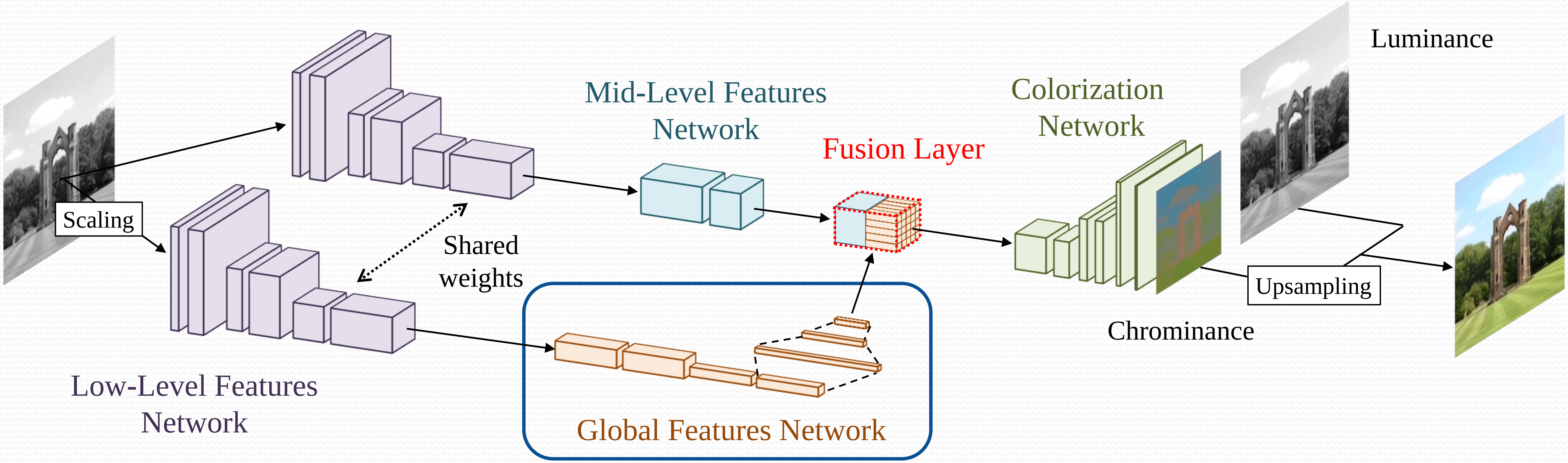
- Two branches: local features and global features
- Composed of four networks

Low-Level Features Network



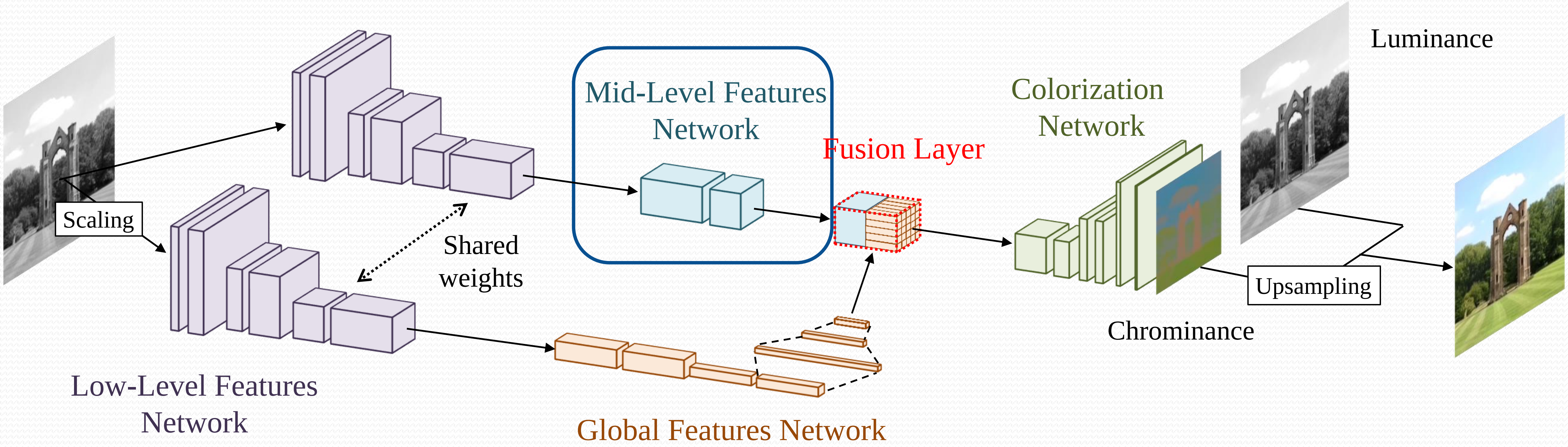
- Extract low-level features such as edges and corners
- Lower resolution for efficient processing

Global Features Network



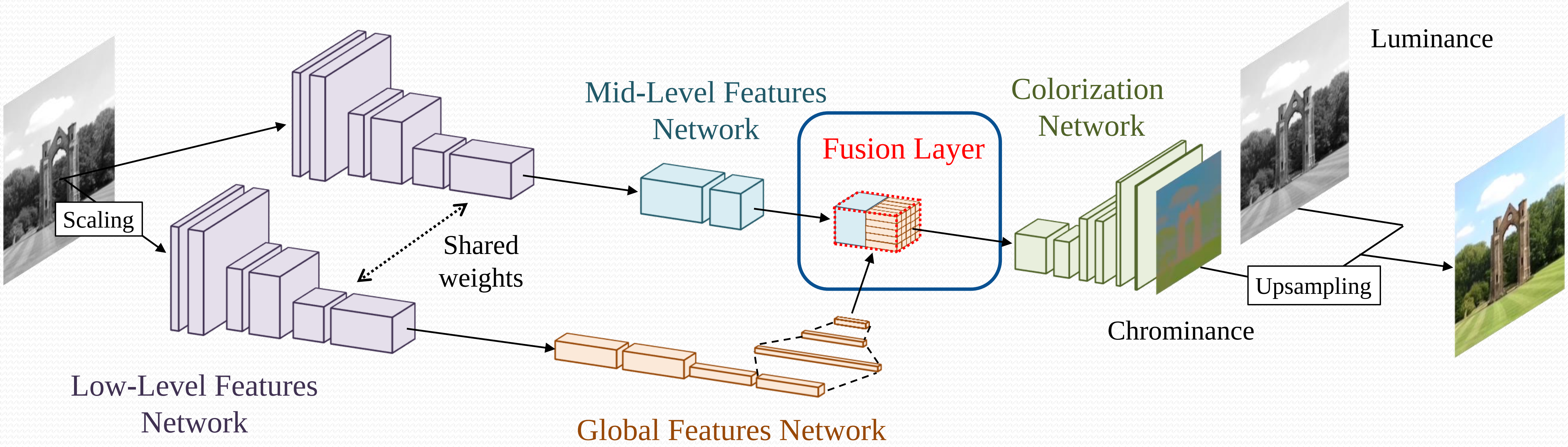
- Compute a **global** 256-dimensional vector representation of the image

Mid-Level Features Network



- Extract mid-level features such as texture

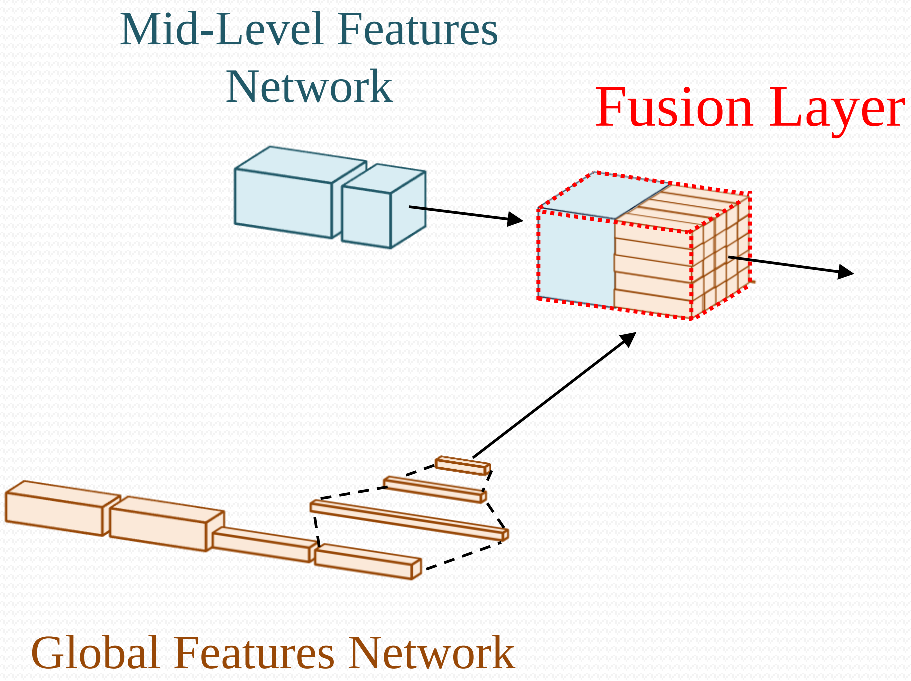
Fusion Layer



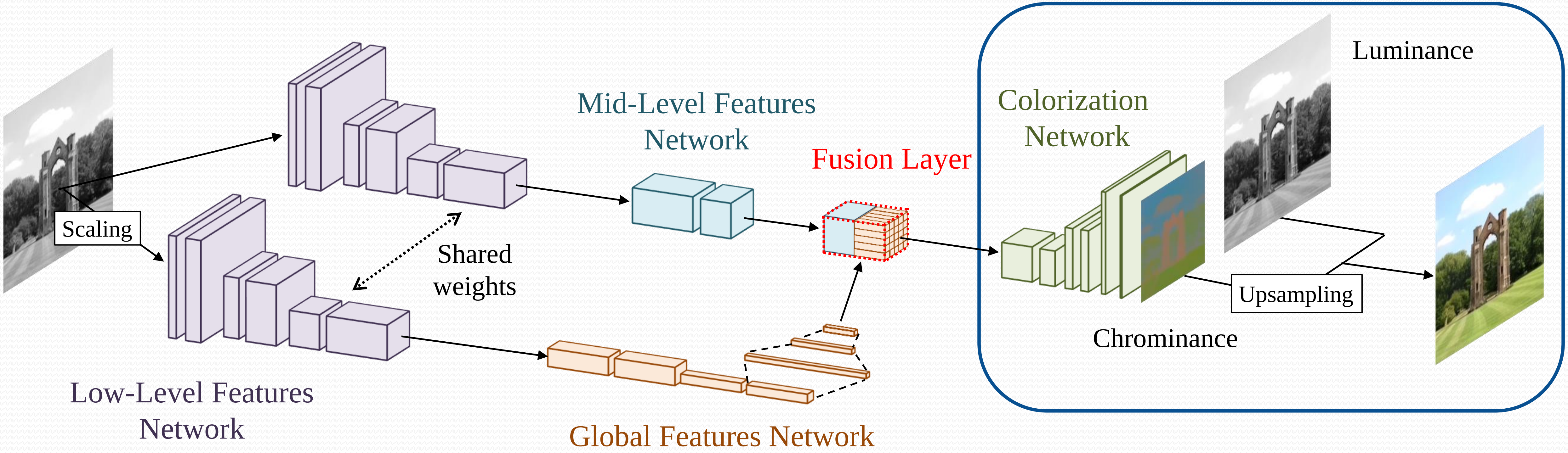
Fusion Layer

- Combine the global features with the mid-level features
- The resulting features are independent of any resolution

$$\mathbf{y}_{u,v}^{\text{fusion}} = \sigma\left(\mathbf{b} + W\begin{bmatrix}\mathbf{y}^{\text{global}} \\ \mathbf{y}_{u,v}^{\text{mid}}\end{bmatrix}\right)$$



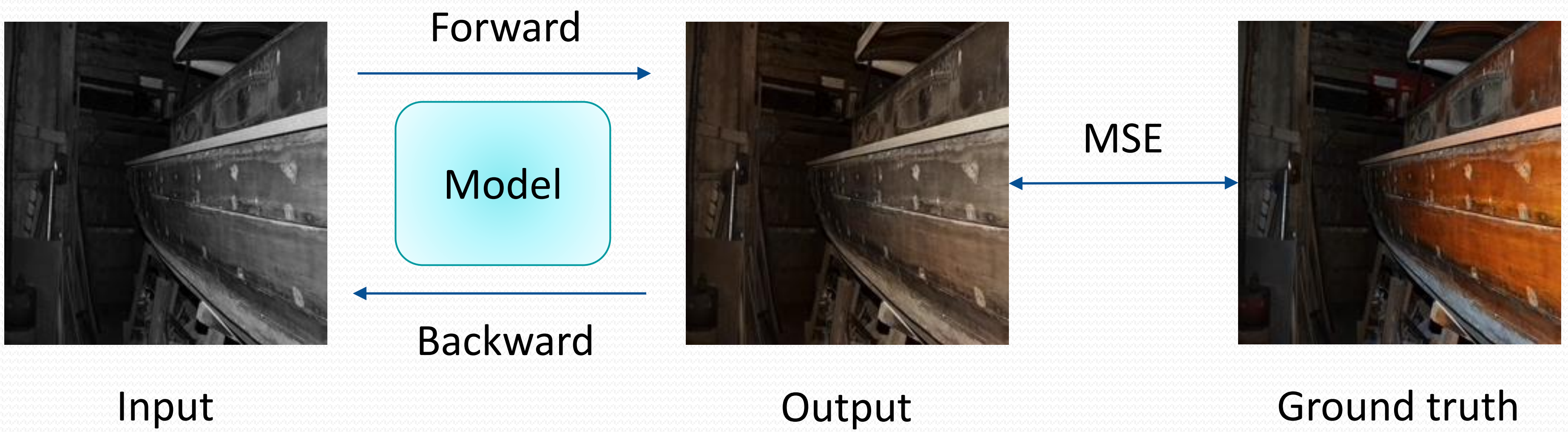
Colorization Network



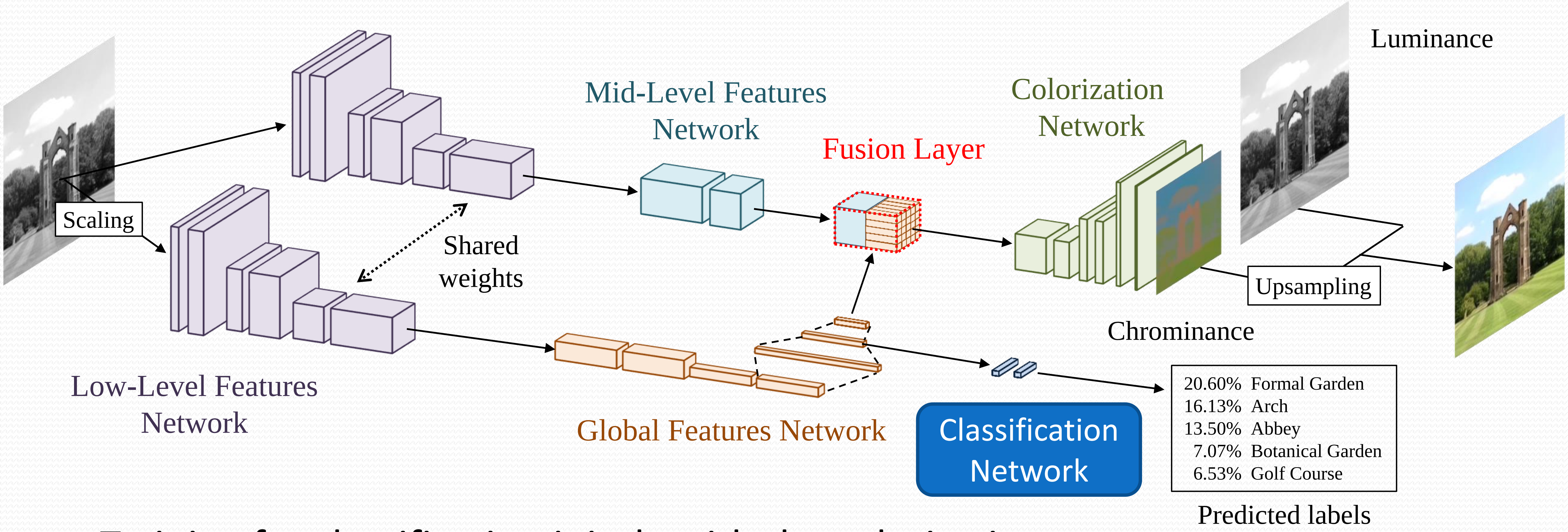
- Compute chrominance from the fused features
- Restore the image to the input resolution

Training of Colors

- Mean Squared Error (MSE) as loss function
- Optimization using ADADELTA [Zeiler 2012]
  - Adaptively sets a learning rate



Joint Training



- Training for classification jointly with the colorization
  - Classification network connected to the global features

Dataset

- MIT Places Scene Dataset [Zhou+ 2014]
- 2.3 million training images with 205 scene labels
  - 256 × 256 pixels


Abbey


Airport terminal


Aquarium


Baseball field

…


Dining room


Forest road


Gas station


Gift shop

…

Results

Computational Time

- Colorize within a few seconds

Image Size	Pixels	CPU (s)	GPU (s)	Speedup
$224 \times 224^{\dagger}$	50,176	0.399	0.080	$5.0\times$
512×512	262,144	1.676	0.339	$4.9\times$
1024×1024	1,048,576	5.629	1.084	$5.2\times$
2048×2048	4,194,304	20.116	4.218	$4.8\times$

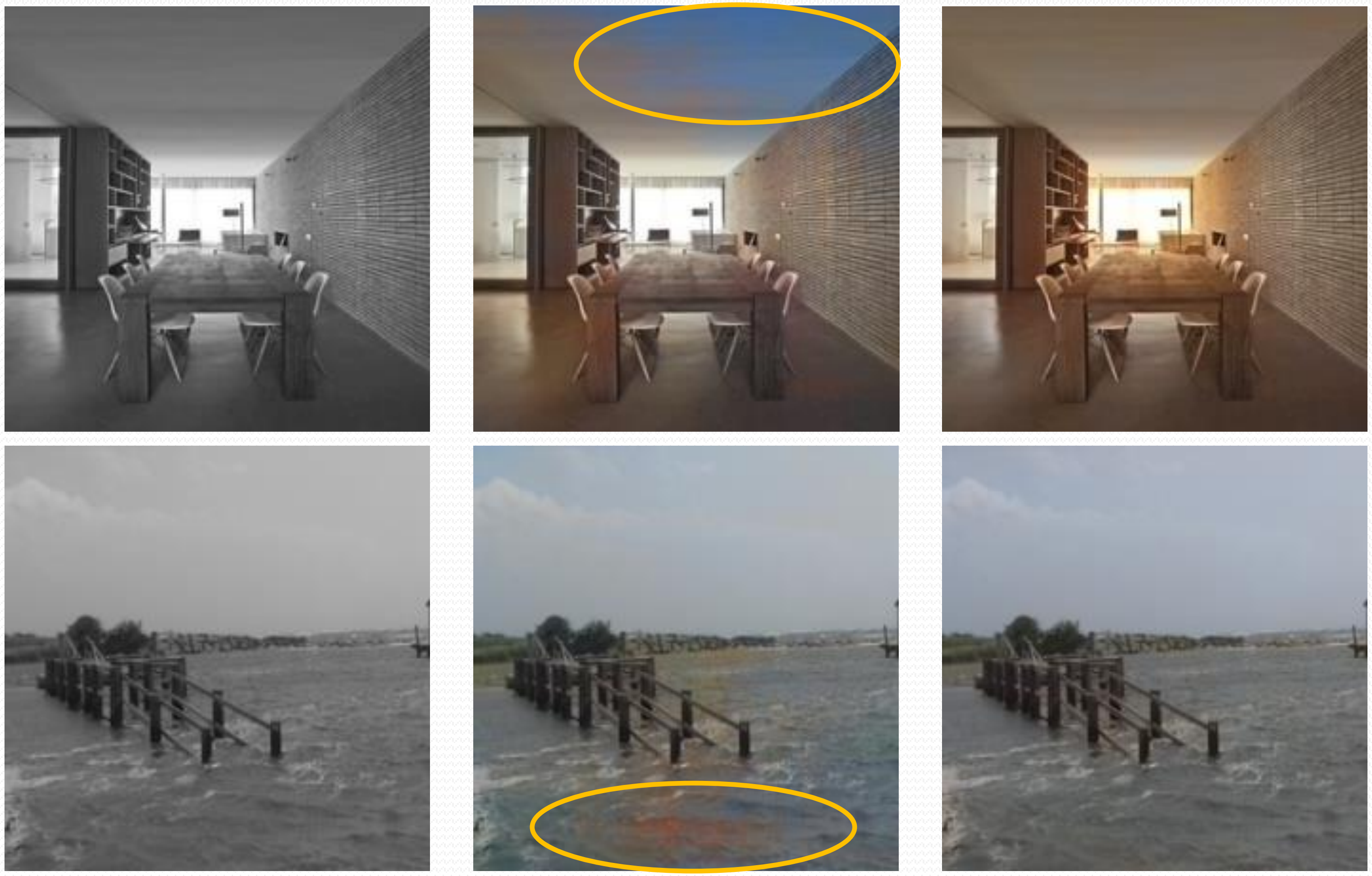
80ms

Colorization of MIT Places Dataset

Comparisons

Input | [Cheng+ 2015] | Ours (w/o global features) | Ours (w/ global features)

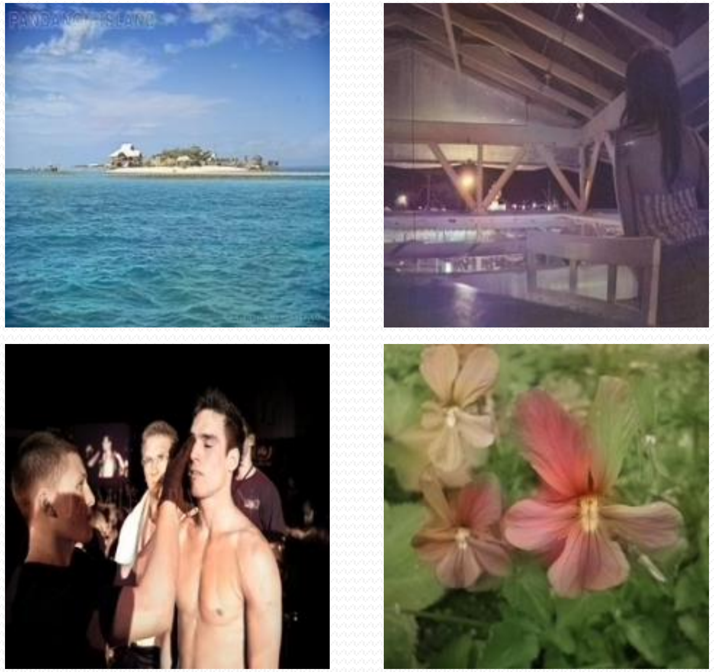
Effectiveness of Global Features



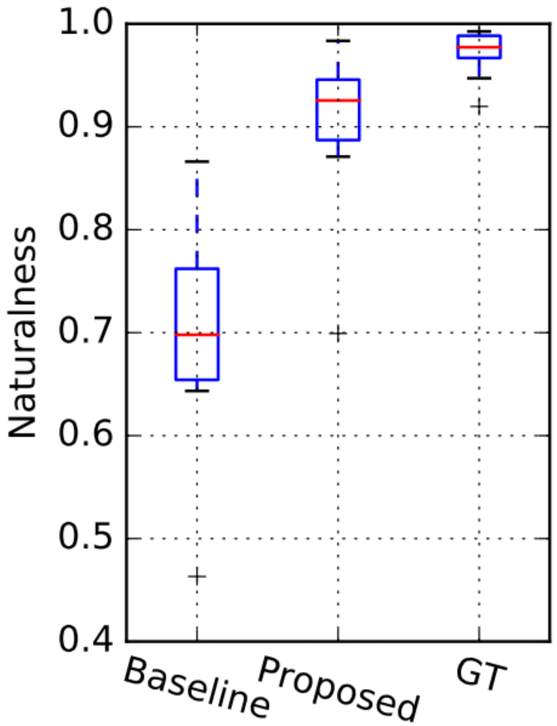
Input w/o global features w/ global features

User Study

- 10 users participated
- We show 500 images of each type: total 1,500 images per user
- 90% of our results are considered "natural"

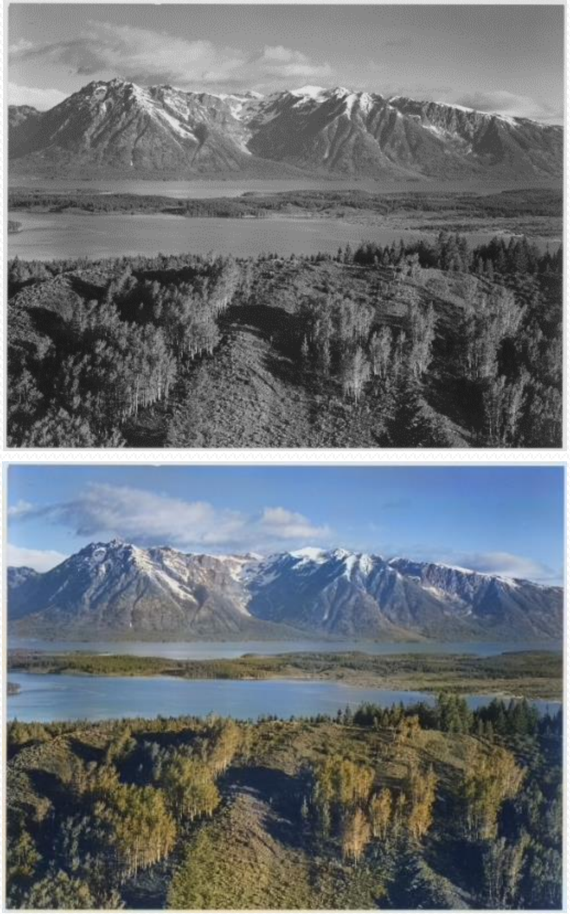


Natural Unnatural



Approach	Naturalness (median)
Ground Truth	97.7%
Proposed	92.6%
Baseline	69.8%

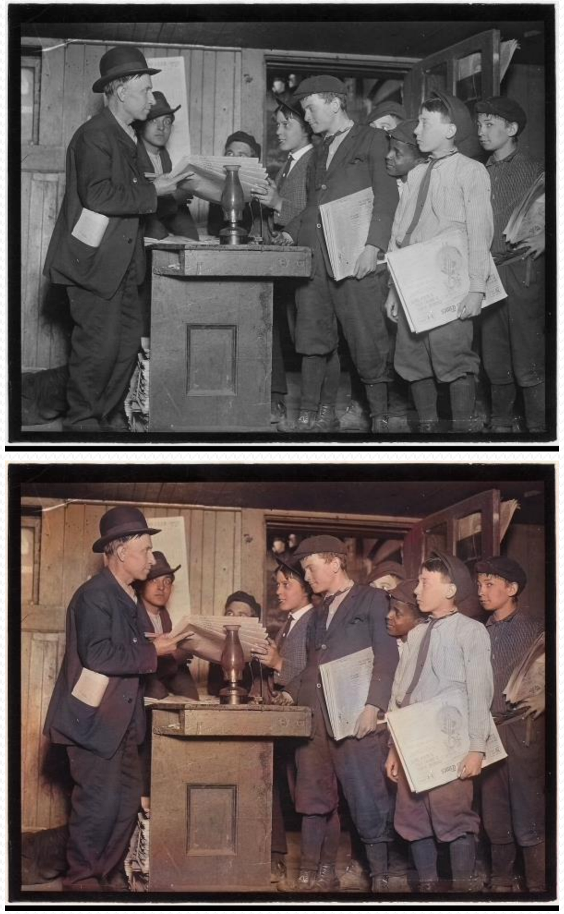
Colorization of Historical Photographs



Mount Moran, 1941

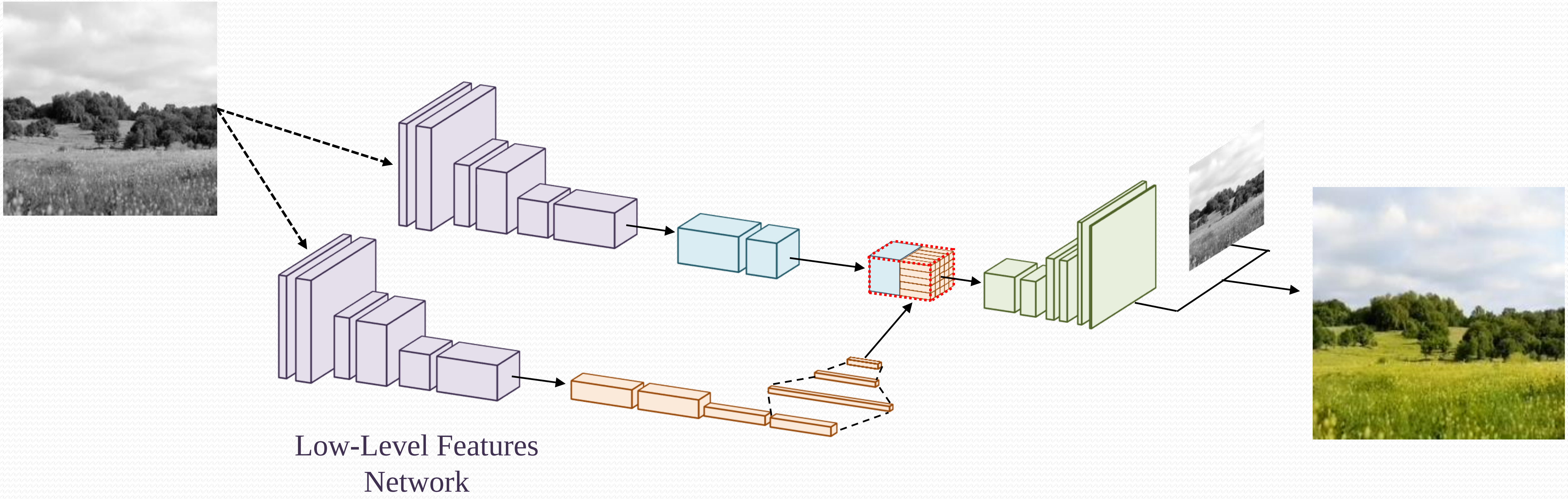
Scott's Run, 1937

Youngsters, 1912

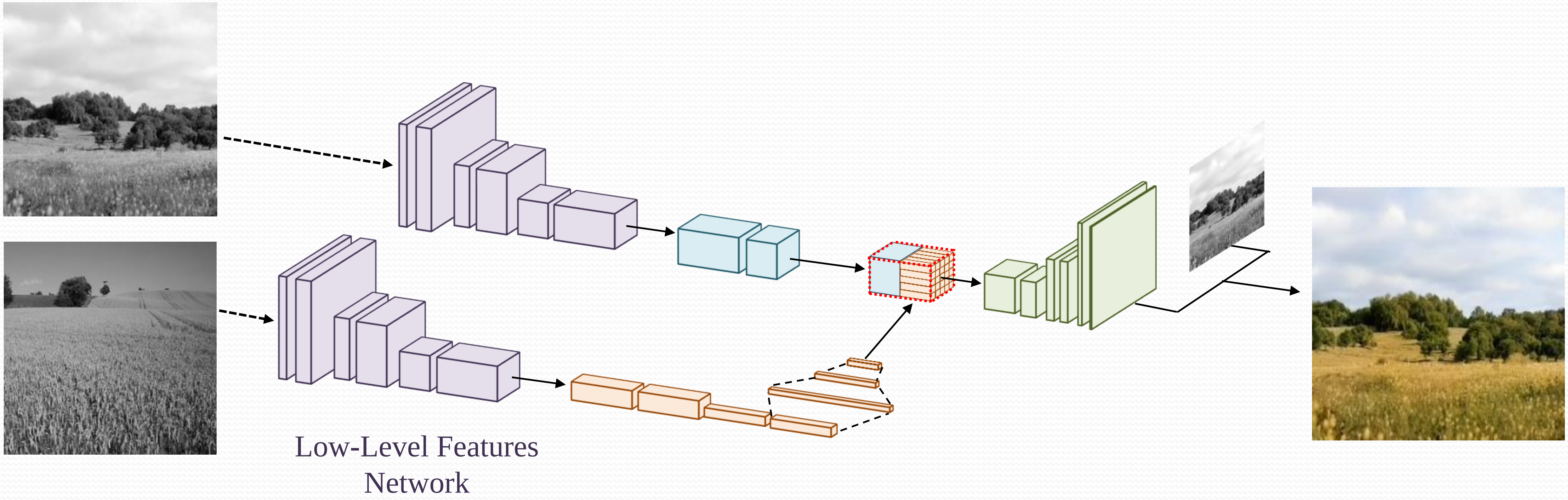


Burns Basement, 1910

Style Transfer



Style Transfer



Style Transfer

- Adapting the colorization of one image to the style of another

Limitations

- Difficult to output colorful images

Input

Ground truth

Output

- Cannot restore exact colors

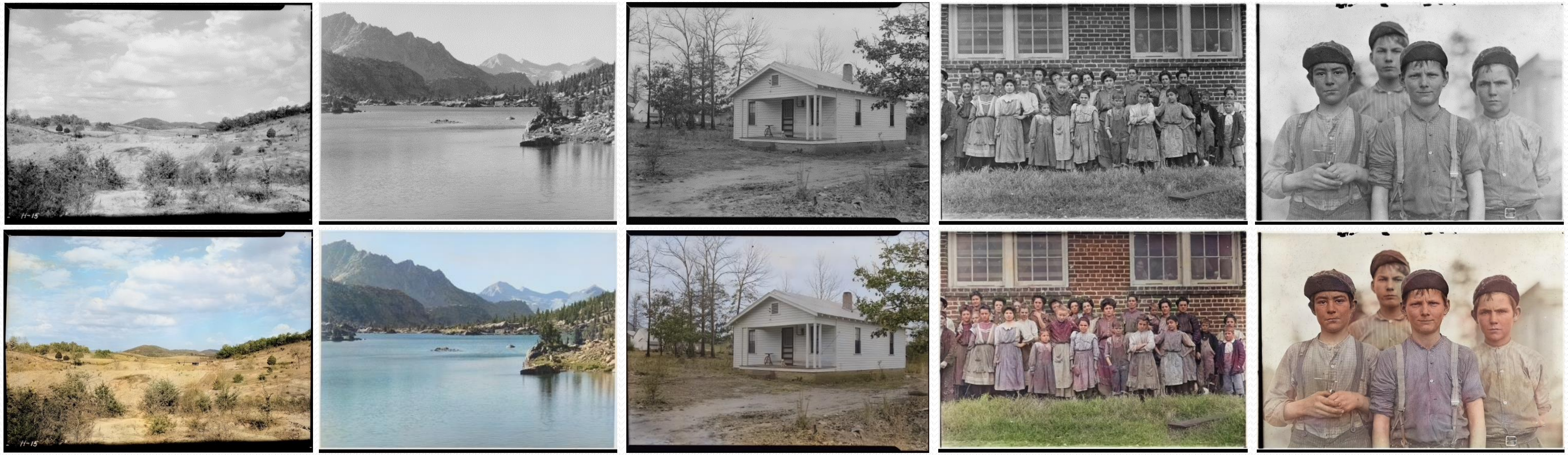
Input

Ground truth

Output

Conclusion

- Novel approach for image colorization by fusing global and local information
  - Fusion layer
  - Joint training of colorization and classification
  - Style transfer



Farm Land, 1933 | California National Park, 1936 | Homes, 1936 | Spinners, 1910 | Doffer Boys, 1909

Thank you!

- Project Page http://hi.cs.waseda.ac.jp/~iizuka/projects/colorization
- Code on GitHub! https://github.com/satoshiiizuka/siggraph2016_colorization

Community Center, 1936

North Dome, 1936

Norris Dam, 1933

Miner, 1937