BMVC 2017 LONDON

PIC: Probabilistic Image Colorization

Amélie Royer, Alexander Kolesnikov, Christoph H. Lampert

IST Austria (Institute of Science and Technology Austria), Klosterneuburg, Austria

IST AUSTRIA Institute of Science and Technology

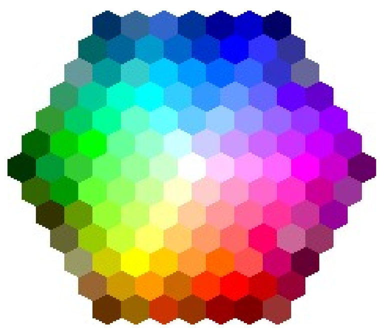
erc European Research Council

Motivation

Image colorization

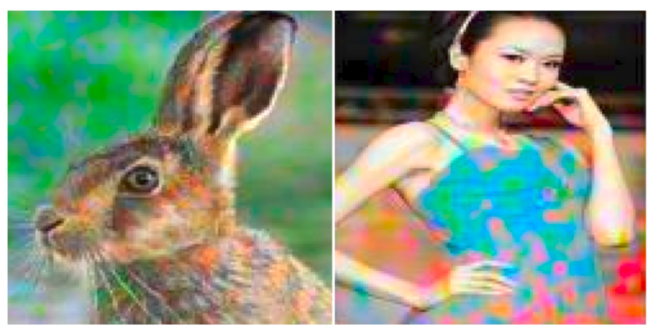
- Grayscale image $\longrightarrow$ colored image
- Multiple plausible "ground-truth" results

Current shortcomings



Discrete colorspace

Desaturated samples



Pixelwise sampling

Proposed model

- Exploit a meaningful feed-forward embedding
- Small autoregressive generative component
- Proper probabilistic framework for sampling
- Model likelihood as a quantitative metric

Background

Colorization task

- **Input**: Grayscale image $\boldsymbol{X^L}$ (luminance channel)
- **Output**: Distribution over the ***ab*** chrominance, $\boldsymbol{p(X^{ab}|X^L)}$

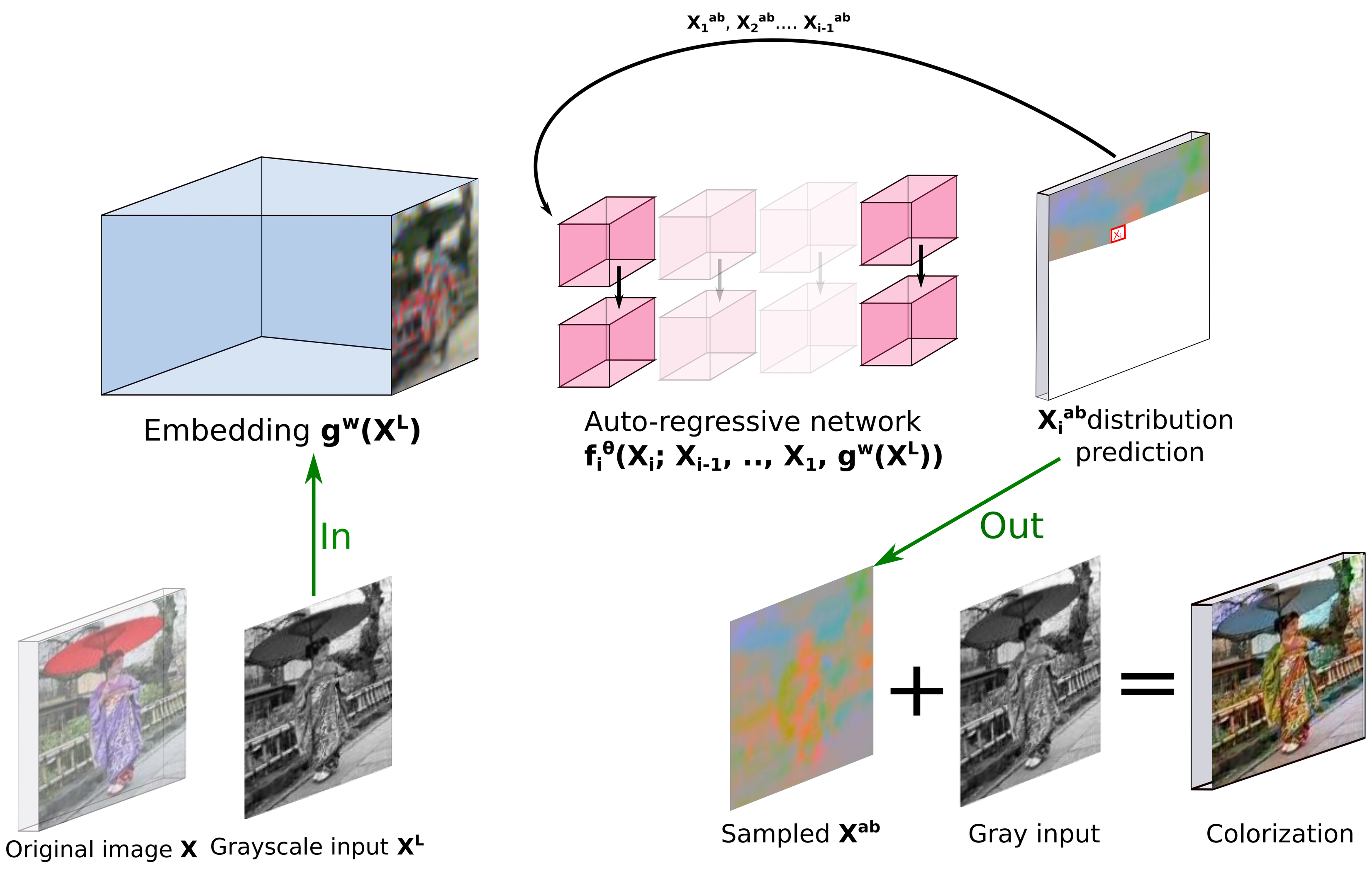
Main issues

- Continuous output space
- One-to-many problem
- No consensus on a quantitative metric

Related Work

- Colorization with a feed-forward CNN **[1]**
- Autoregressive generative models **[2]**

Probabilistic Image Colorization Workflow



Autoregressive probabilistic models

Model. We use the chain rule to decompose the probability distribution over pixels in the colorized image:

$$p(X^{ab}|X^L) = \prod_{i=1}^{N} p(X_i^{ab}|X_{1\dots i-1}^{ab}; X^L)$$

Sampling. Sampling is done iteratively, pixel by pixel (raster order) starting from $\widehat{X}_1^{ab} \sim p(X_1^{ab}|X^L)$ and then

$$\forall i,\ \widehat{X}_i^{ab} \sim p(X_i^{ab}|\widehat{X}_{1\dots i-1}^{ab}; X^L)$$

Network architecture

$$p(X_i^{ab}|X_{1\dots i-1}^{ab}; X^L) = f_i^\theta(X_{1\dots i-1}^{ab}; g^w(X^L)), \text{ where}$$

g^w is a conditioning CNN which outputs an embedding of X^L

f^θ is an autoregressive network, which outputs a multimodal discrete distribution over the colorspace

[1] R. Zhang, P. Isola, and A. Efros, *"Colorful image colorization"*, ECCV'16
[2] A. van den Oord, N. Kalchbrenner, and K. Kavukcuoglu, *"Pixel recurrent neural networks"*, ICML'17.
[3] G. Larsson, M. Maire, and G. Shakhnarovich, *"Learning representations for automatic colorization"*, ECCV'16
[4] S. Iizuka, E. Simo-Serra, and H. Ishikawa, *"Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification"*, SIGGRAPH'16
[5] S. Guadarrama, R. Dahl, D. Bieber, M. Norouzi, J. Shlens, K. Murphy, *"PixColor: Pixel Recursive Colorization"*, BMVC'17

Training procedure

Model likelihood as quantitative metric

We use the model likelihood $\boldsymbol{p(X^{ab}|X^L)}$ during training (maximization) and for evaluation (model selection)

Chrominance distribution

- We model the *ab*-color distribution as a mixture of 10 logistic components, which are parametrized by 100 outputs
- We sample $\boldsymbol{X^{ab}}$ at one-fourth resolution and upscale it for the colorized image

Architecture

- g^w is a feed-forward network with 30 convolutional layers
- f^θ is a conditional PixelCNN++ architecture with 8 residual blocks

Qualitative results on ILSVRC2012

PIC produces colorful and diverse samples

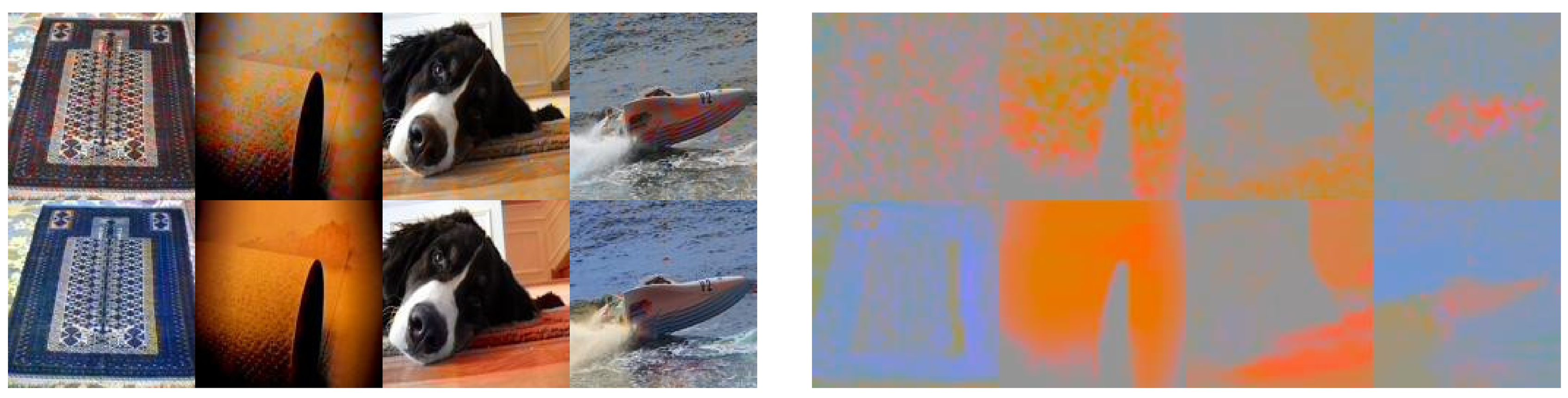
Unambiguous objects' colors are consistent (e.g., sky, grass)

Failure cases

PIC might fail to capture long-range pixel interactions in complex scenes

Additional experiments

Ablation: Autoregressive component



PIC Likelihood: **2.51**, CNN Likelihood: **4.53**

Comparison to baselines

Conclusions

[+] Rigorous probabilistic framework for colorization
[+] Likelihood serves as a principled evaluation metric
[+] No ad-hoc heuristics are required

[-] Linear chrominance upscaling [5]
[-] Qualitative results can be improved