Recovering realistic texture in image super-resolution by deep spatial feature transform

Xintao Wang Ke Yu Chao Dong Liang Lin Chen Change Loy

Issues

1. How to represent the semantic categorical prior?

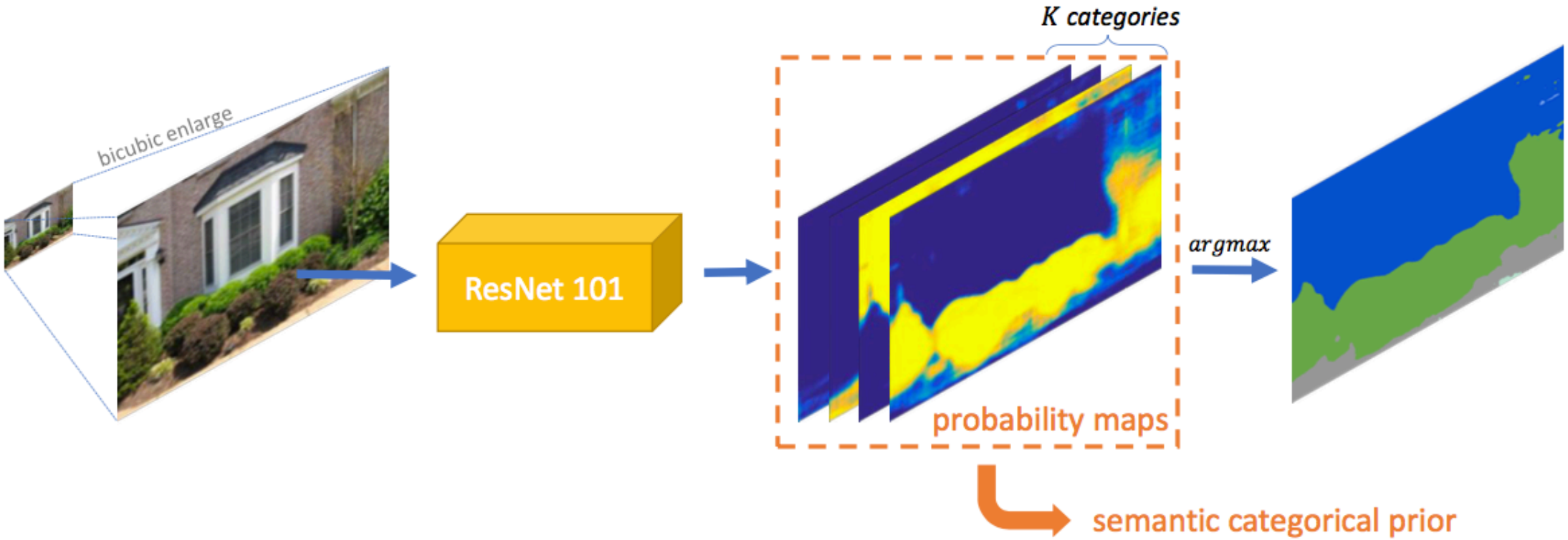
   ❑ Our approach:
   explore semantic segmentation probability maps as the categorical prior up to pixel level.

2. How categorical prior can be incorporated into the reconstruction process effectively?

   ❑ Our approach:
   propose a novel **S**patial **F**eature **T**ransform that is capable of altering the network behavior conditioned on other information.

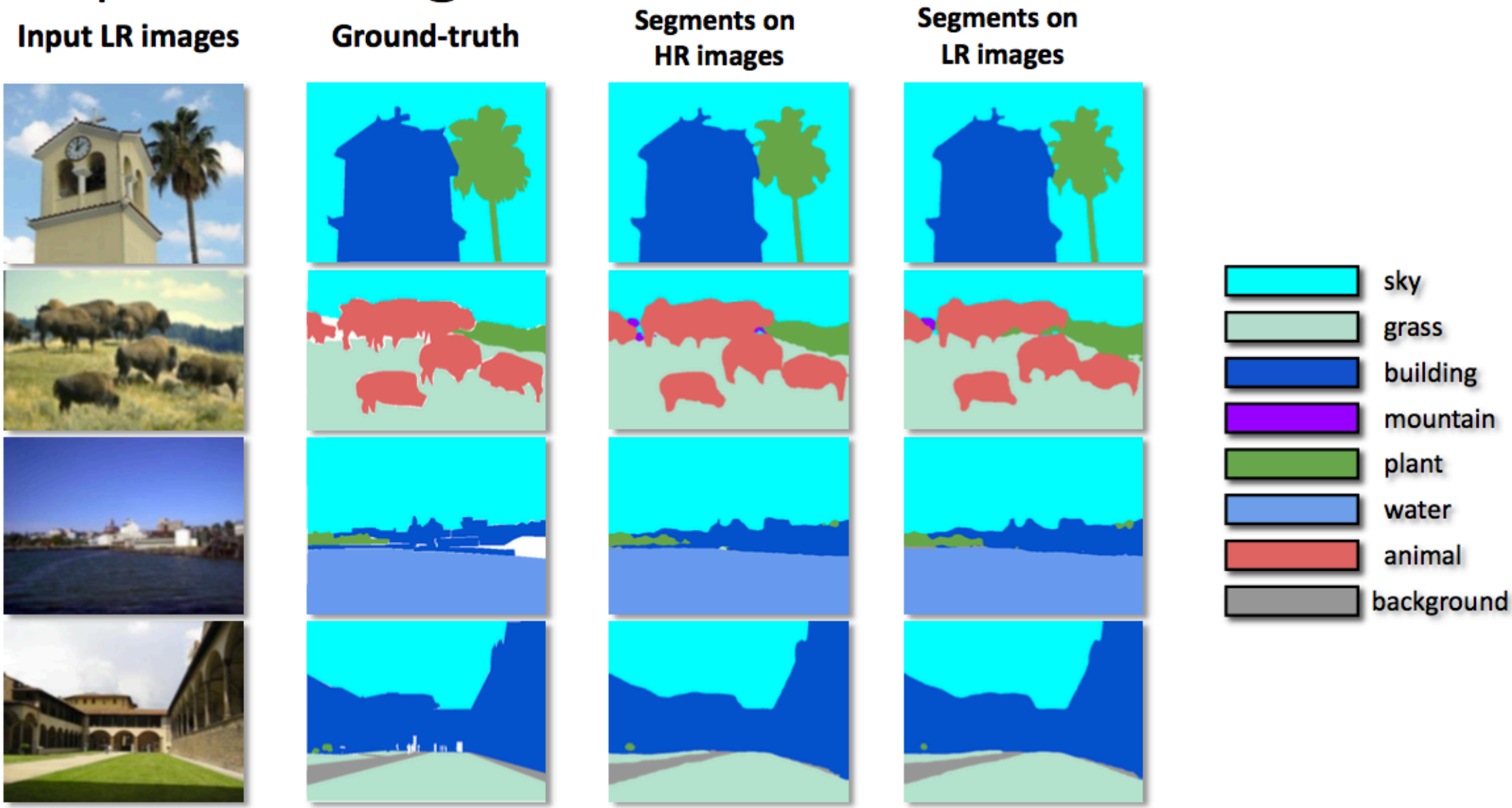
Represent categorical prior

- **Contemporary CNN segmentation network[1]**
  - fine-tuned on LR images



[1]Z. Liu, X. Li, P. Luo, C.-C. Loy, and X. Tang. Semantic image segmentation via deep parsing network. In ICCV, 2015
http://ice.dlut.edu.cn/valse2018/ppt/2018ValseDChao.pdf

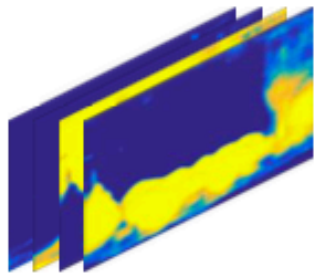
Examples on segmentation



http://ice.dlut.edu.cn/valse2018/ppt/2018ValseDChao.pdf

Incorporate conditions

Categorical prior



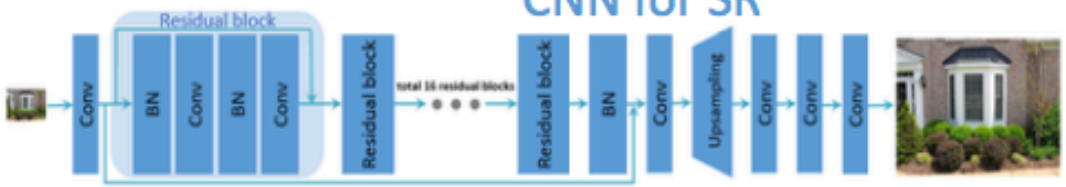
prior Ψ

probability maps $P_1, P_2, \ldots, P_K$

$$\Psi = (P_1, P_2, \ldots, P_K)$$

$$\hat{\boldsymbol{y}} = G_\theta(\boldsymbol{x}|\Psi)$$

CNN for SR



input LR image $\boldsymbol{x}$

net G parametrized by θ

restored image $\hat{\boldsymbol{y}}$

$$\hat{\boldsymbol{y}} = G_\theta(\boldsymbol{x})$$

http://ice.dlut.edu.cn/valse2018/ppt/2018ValseDChao.pdf

Spatial Feature Transform

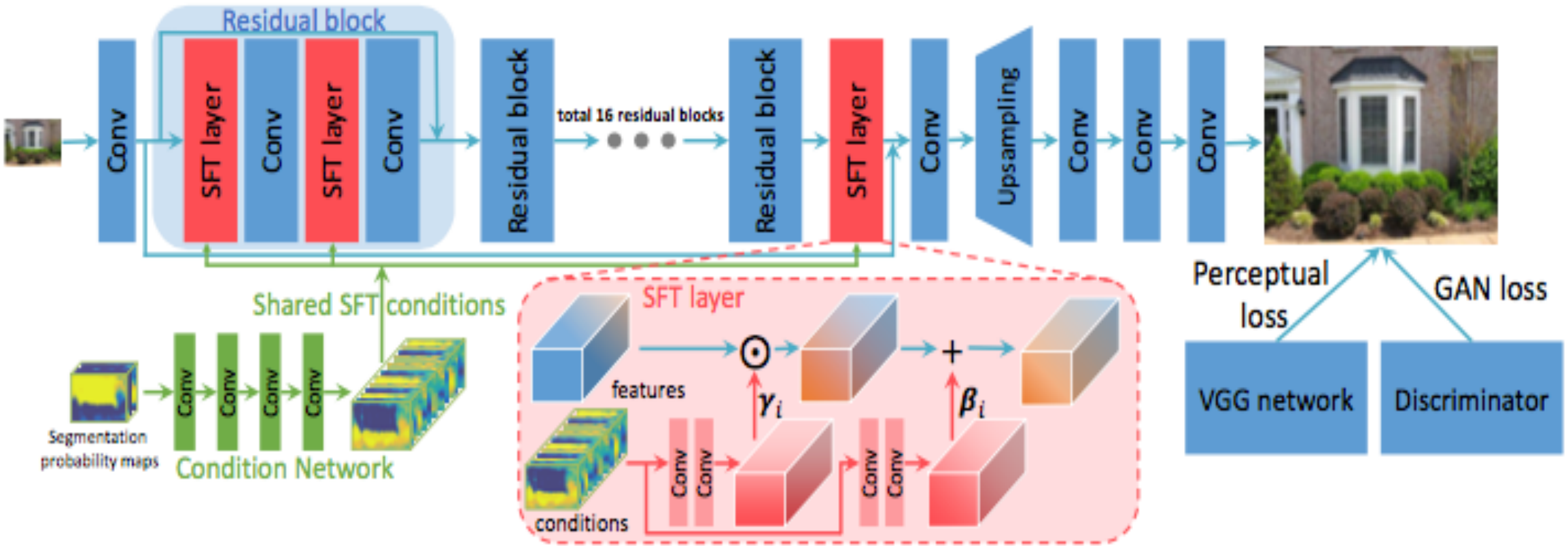
- By learning a mapping function $\mathcal{M}$, the prior Ψ is modeled by a pair of affine transformation parameters (γ, β) .

$$\mathcal{M}: \Psi \mapsto (\gamma, \beta)$$

- The modulation is then carried out by an affine transformation on feature maps $\boldsymbol{F}$. $\mathrm{SFT}(\boldsymbol{F}|\gamma, \beta) = \gamma \odot \boldsymbol{F} + \beta$

$$\hat{\boldsymbol{y}} = G_\theta(\boldsymbol{x}|\Psi) \xrightarrow[\mathrm{SFT}(\boldsymbol{F}|\gamma,\beta) = \gamma \odot \boldsymbol{F} + \beta]{\mathcal{M}: \Psi \mapsto (\gamma, \beta)} \hat{\boldsymbol{y}} = G_\theta(\boldsymbol{x}|\gamma, \beta)$$

Spatial Feature Transform

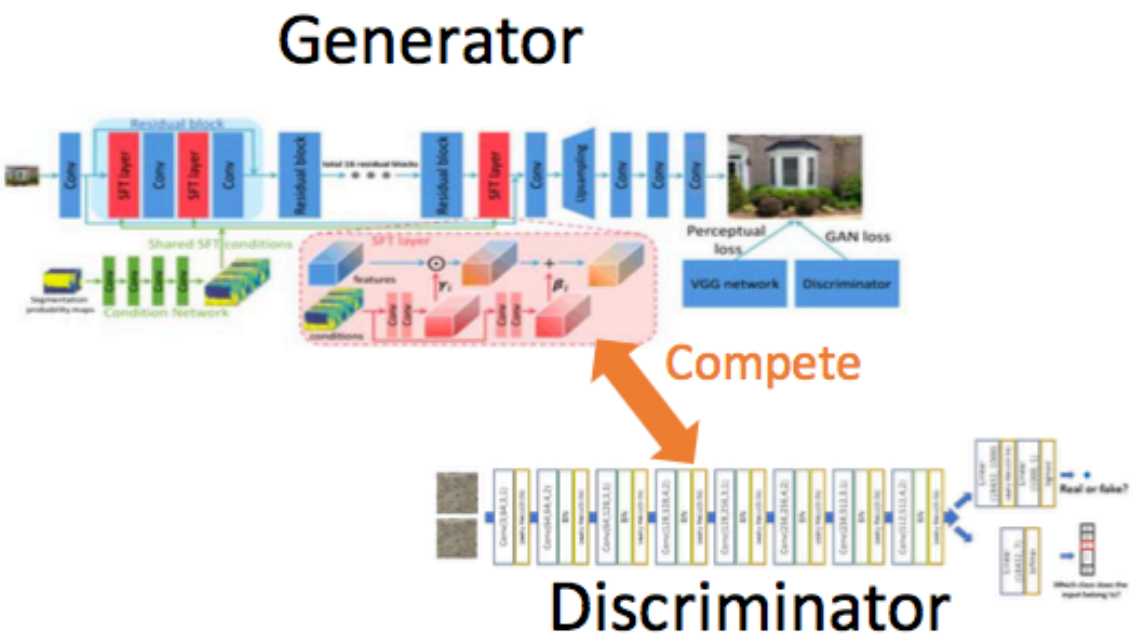


http://ice.dlut.edu.cn/valse2018/ppt/2018ValseDChao.pdf

loss function

- Adversarial loss[1]
  - ✓ encourage the network to generate images that reside on the manifold of natural images

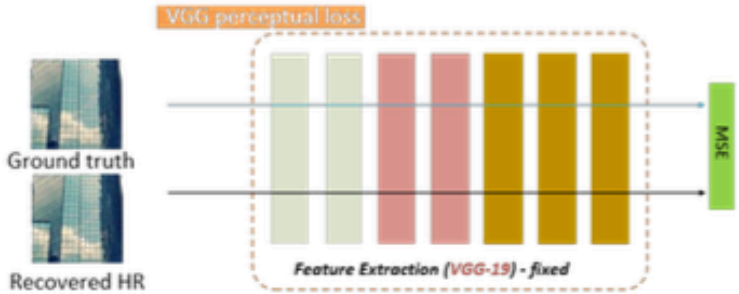
$$\min_{\theta}\max_{\eta} \mathrm{E}_{y\sim p\mathrm{HR}}\, logD_{\eta}(y) + \mathrm{E}_{x\sim p\mathrm{LR}}\log(1 - D_{\eta}(G_{\theta}(x)))$$



- Perceptual loss[2]

  use a pre-trained 19-layer VGG network (features before conv54)
  - ✓ optimize a super-resolution model in a feature space

$$\|\phi_{VGG}(\hat{y}) - \phi_{VGG}(y)\|_2^2$$



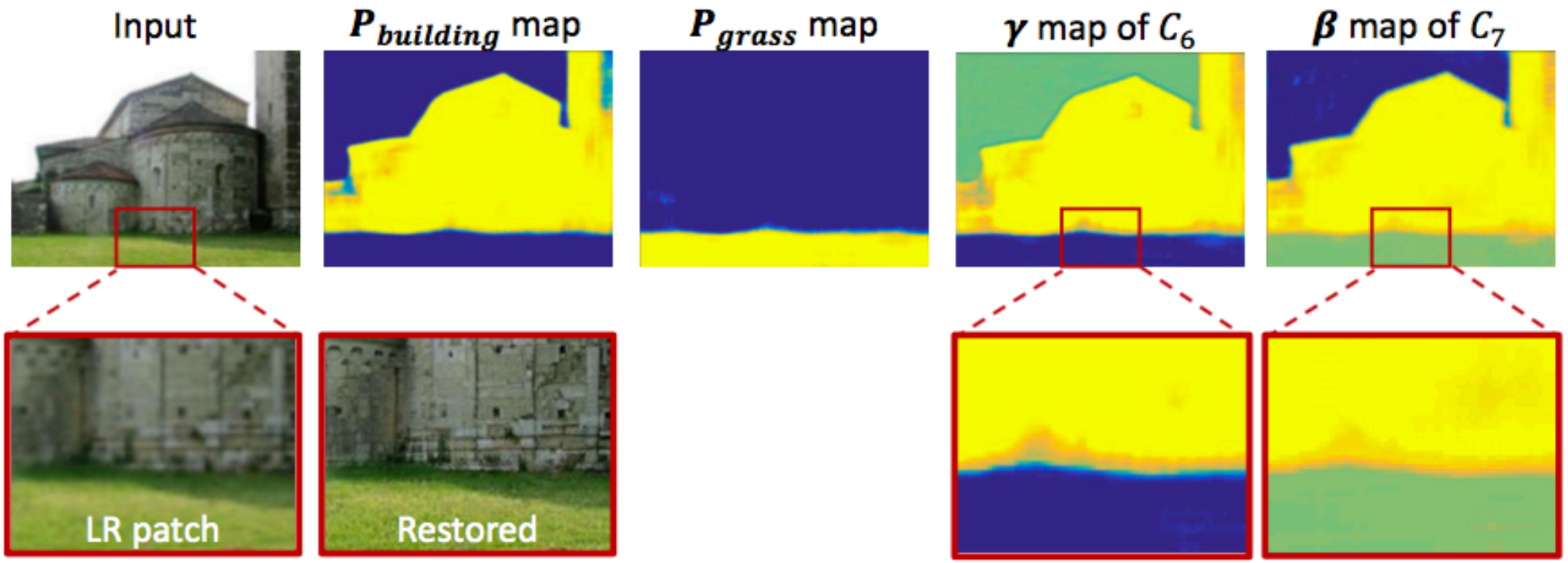
[1]Goodfelow, lan, et al. Generative adversarial nets. In NIPS. 2014

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

http://ice.dlut.edu.cn/valse2018/ppt/2018ValseDChao.pdf

Spatial condition

- The modulation parameters (γ, β) have a close relationship with probability maps $\boldsymbol{P}$ and contain spatial information.



http://ice.dlut.edu.cn/valse2018/ppt/2018ValseDChao.pdf

Results

http://ice.dlut.edu.cn/valse2018/ppt/2018ValseDChao.pdf

Conclusion

- Explore semantic segmentation maps as categorical prior for realistic texture recovery.
- Propose a novel Spatial Feature Transform layer to efficiently incorporate the categorical conditions into a CNN-based SR network.
- Extensive comparisons and a user study demonstrate the capability of SFT-Net in generating realistic and visually pleasing textures.