Modeling Deformable Gradient Compositions for Single-Image Super-resolution

Yu Zhu[1], Yanning Zhang[1], Boyan Bonev[2], Alan L. Yuille[2]
[1]School of Computer Science, Northwestern Polytechnical University, [2]Department of Statistics, University of California, Los Angeles

Single-image super-resolution is becoming more important with the development of high-definition display devices. However, recovering the high-resolution (HR) details from single low-resolution (LR) image is still challenging. For example-based methods, the ambiguity of HR/LR patch pairs is a common problem. We observe that singular structures such as single edges and corners are more robust to scale change, as pointed out in some works [3, 7]. In other words they are less ambiguous across different scales. Meanwhile, most methods work better on singular structures. This inspires us to decompose the non-singular structures to single ones. We also exploit the approach of [9] to make the dictionary deformable and more expressive.

This paper proposes a novel deformable compositional model for single-image super-resolution. Both the patch in the LR input image and the dictionary patch are decomposed to singular structures by using masks. For each input LR patch containing a singular structure, its best match in the dictionary is deformed to recover the gradient field. Finally the HR gradient information is integrated into the LR input image.

We start from patches centered at the gradient ridge points instead of following the raster-scan strategy. A gradient ridge point is the local maximum along the gradient descending path. For each gradient ridge position $p=(x,y)$, we first extract the patches centering at (x,y) by $\boldsymbol{R}_p$ and integrate it with the mask $M_{p,Z}$. Given the LR dictionary D_l and the LR gradient dictionary $D_{gl}=(D_{xl},D_{yl})$, we choose the best element from the corresponding HR dictionary D_h by function $\boldsymbol{\delta}$. After the deformation ϕ and the contrast adjustment by η_p, $\boldsymbol{R}_p^\top$ maps the patch back to the position p within the reconstructed image. $\sum_{p\in Z}\boldsymbol{R}_p^\top M_{p,X_l}^\top M_{p,X_l}\boldsymbol{R}_p$ counts the masks accumulated on each pixel. In other words, our method averages all the overlapping masked dictionary elements. The predicted HR gradient $\nabla\tilde{X}_h$ is reconstructed as the following equation:

$$\nabla\tilde{X}_h=\frac{\sum_{p\in Z}\boldsymbol{R}_p^\top M_{p,Z}^\top\cdot\eta_p\cdot\phi(D_h\boldsymbol{\delta}(D_{gl},M_{p,Z}\boldsymbol{R}_p\nabla X_l))}{\sum_{p\in Z}\boldsymbol{R}_p^\top M_{p,Z}^\top M_{p,Z}\boldsymbol{R}_p} \quad (1)$$

The symbols are the following:

Z: the gradient ridges of the entire image *i.e.* the local maximum along the gradient direction (See Section 3.2).

$\boldsymbol{R}_p$: patch extraction operator that extracts a patch centered at the position $p=(x,y)$.

$\boldsymbol{R}_p^\top$: the inverse operation of $\boldsymbol{R}_p$ that maps the patch to the position p of the constructed image.

$M_{p,Z}$: the patch mask generated from the gradient ridge point, centered at p (See Section 3.2).

$\boldsymbol{\delta}$: the indicative function that chooses the best match between the input LR patch and LR gradient dictionary (See Section 3.3 and 3.4).

ϕ: the deformation function elaborated in Section 3.5.

η_p: the gradient contrast adjustment ratio between the LR patch and corresponding HR patch with the form $\eta_p=\alpha Var(M_{p,Z}\boldsymbol{R}_p|\nabla X_l|)$, where $|\nabla X_l|$ denotes the gradient magnitude. By the global parameter α, η_p we adjust the contrast of the normalized output of the patch deformation stage. The setting of the global parameter α is discussed in the experiments.

D_l and D_h: the LR and HR dictionary respectively. Note that we use masked patches in Eqn. (1), even though we only display the raw gradient dictionary elements and the masks separately in Figure 1.

After the estimated image gradient $\nabla\tilde{X}_h$ is obtained ($\nabla_x\tilde{X}_h$ and $\nabla_y\tilde{X}_h$ are processed separately), we impose the integrated gradient prior $\nabla\tilde{X}_h$ on the given LR image X_l to recover the HR details. The same strategy is employed by the work [4] The following energy function is minimized by enforcing the constraint in both intensity domain and gradient domain:

$$E(X_h|X_l,\nabla\tilde{X}_h)=||SHX_h-X_l||^2+\beta||\nabla X_h-\nabla\tilde{X}_h||^2, \quad (2)$$

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

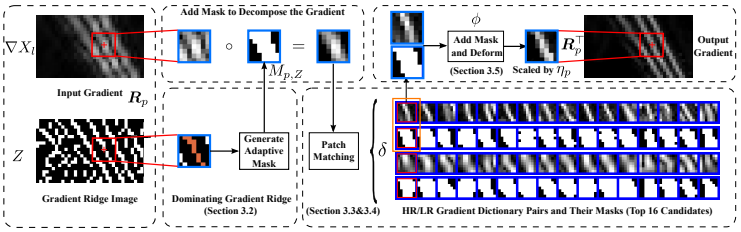

Figure 1: Flowchart of the proposed method

where S is a down-sampling operator, H is a blurring operator. β is a parameter that balances the constraints between the intensity domain and the gradient domain. The global minimum can be obtained by gradient decent:

$$X_h^{t+1}=X_h^t-\tau[(H^\top S^\top SHX_h-H^\top S^\top X_l)-\beta(\mathrm{div}(\nabla X_h)-\mathrm{div}(\nabla\tilde{X}_h))] \quad (3)$$

where t is the iteration number and τ is the iteration step. $\mathrm{div}(\nabla X_h)$ denotes the divergence of ∇X_h via the form $\mathrm{div}(\nabla X_h)=\partial^2X_h/\partial x^2+\partial^2X_h/\partial y^2$, which can be implemented easily using the Laplace operator.

Our performance evaluation is based on the image test Set 5 and Set 14. These images are also the main test sets in the literature [2, 5, 8]. Average PSNR/SSIM performance on luminance channel are evaluated as Table 1:

Table 1: Average performance in PSNR and SSIM on the Set 5 and 14 (3×)

Average	SCSR[6]	DPSR[9]	SRCNN[2]	DNC[1]	Proposed
PSNR	29.63	29.70	29.86	29.92	**30.07**
SSIM	0.8899	0.8908	0.8901	0.8927	**0.8952**

In this paper, we propose a Deformable Gradient Compositional model to represent the non-singular structures as compositions of single ones, each of which is allowed some deformation. In our future work we plan to address the decomposition problem for the joint edges or T-junctions which is not explicitly handled in our work.

The authors would like to thank Liang-Chieh Chen for the useful comments. The work is supported by Chinese Scholarship Council and grants NSF of China (61231016, 61301193, 61303123, 61301192), NPU-FFR-JCT20130109, ONR MURI N000014-10-1-0933 and NIH 5R01EY022247-03.

[1] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen. Deep network cascade for image super-resolution. In *ECCV*, 2014.

[2] C. Dong, C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014.

[3] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *ICCV*, pages 349–356, 2009.

[4] J. Sun, J. Sun, Z. Xu, and H. Shum. Gradient profile prior and its applications in image super-resolution and enhancement. *TIP*, 20(6):1529–1542, June 2011.

[5] R. Timofte, V. De, and L. Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *ICCV*, pages 1920–1927, Dec 2013.

[6] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE TIP*, 19(11):2861–2873, 2010.

[7] J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. In *CVPR*, pages 1059–1066, 2013.

[8] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *International Conference on Curves and Surfaces*, pages 711–730, 2010.

[9] Y. Zhu, Y. Zhang, and A. Yuille. Single image super-resolution using deformable patches. In *CVPR*, pages 2917–2924, 2014.