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, China
[2]Department of Statistics, University of California, Los Angeles, USA
zhuyu1986@mail.nwpu.edu.cn, ynzhang@nwpu.edu.cn, bonev@ucla.edu, yuille@stat.ucla.edu

Abstract

*We propose a single-image super-resolution method based on the gradient reconstruction. To predict the gradient field, we collect a dictionary of gradient patterns from an external set of images. We observe that there are patches representing singular primitive structures (*e.g. *a single edge), and non-singular ones (*e.g. *a triplet of edges). Based on the fact that singular primitive patches are more invariant to the scale change (*i.e. *have less ambiguity across different scales), we represent the non-singular primitives as compositions of singular ones, each of which is allowed some deformation. Both the input patches and dictionary elements are decomposed to contain only singular primitives. The compositional aspect of the model makes the gradient field more reliable. The deformable aspect makes the dictionary more expressive. As shown in our experimental results, the proposed method outperforms the state-of-the-art methods.*

1. Introduction

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. It is an ill-posed problem because many details are lost during the degradation.

Recent works exploit various priors to regularize the ill-posed problem. Besides the image smoothness prior used by traditional interpolation methods, more sophisticated priors such as gradient based prior [7, 19] are proposed. But modeling the gradient distribution via a few parameters is not easy, because the local image structure is far more complicated than the parametric representation.

Alternatively, example-based methods offer a good way to represent the local structure by patches. These methods either exploit the self-repeated example patches [4, 8, 11] or estimate the high-resolution details from the high/low resolution pairs [2, 4, 9, 13, 15, 21, 24, 25, 26, 28, 30]. Recent work [4, 6] tries to utilize deep networks to learn the inherent relationship between the HR/LR patch pairs.

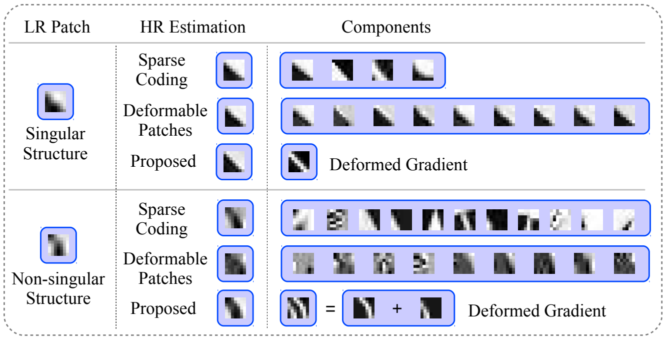


Figure 1. Singular and Non-singular cases approached by sparse coding [24], deformable patches [30] and the proposed method. The HR estimation and the components shown to illustrate that sparse coding and deformable patches work better on singular case than non-singular case. By the proposed deformable compositional model, the non-singular patches can be well reconstructed.

The ambiguity of HR/LR patch pairs is a common problem in the example-based methods. Different HR patches may be corresponded to similar LR patches. To address this problem, Freeman *et al.* [9] consider local consistency between adjacent patches and use Markov Random Fields to find the appropriate candidates from the ambiguous patches. Later methods try to learn the best linear combination of the ambiguous patterns through sparse coding dictionary [24, 25], Deep Network Cascade (DNC) [4] or Convolutions Neural Network(CNN) [6]. Recent work by Zhu *et al.* [30] also suggest that deformable patches can relax the problem.

Singular structures such as single edges and corners are more robust to scale change, as pointed out in some works [1, 11, 23]. In other words they are less ambiguous across different scales. Also, the non-singular structures are more likely to be lost during the degradation. Most methods work better on singular structures. Figure 1 illustrates two broad method categories (sparse coding [24] and deformable patches [30]), showing the HR estimation and their components (linear [25] and weighed [30] combinations). All of them work well on singular structures. But for non-singular case, the sparse coding method tends to select more complex patterns, leading to the blurred HR estimation. And for deformable patches method, it is hard to find appropriate candidates for the non-singular structure. This inspires us to decompose the non-singular structures to singular structures. We also exploit the approach of [30] to make the dictionary deformable and, thus, more expressive. We apply the deformations separately to each component of the compositions. Note that we do not focus on the T-junction structure but on the separable ones *e.g*. the disjoint edges. For simplicity, we treat the non-separable T-junction structures in the same way we treat singular structures. This approximation works well, as the structure is corrected when all the adjacent patches are averaged.

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. In sum, our main contributions are:

1. We propose a deformable compositional model to decompose the non-singular structures into singular structures. Masks are calculated to identify the dominating region of each gradient ridge.

2. We extend the deformable patches based model to the gradient domain, which contributes to the HR reconstruction with sharp edges.

2. Related Work

Single super-resolution methods can be summarized into two categories: the gradient statistical prior based methods [7, 19, 20], and patch example based methods [2, 4, 6, 8, 9, 11, 15, 25, 23, 27, 30].

The gradient statistical prior based methods [7, 19, 20] are parametric methods which try to model the gradient transfer function from LR to HR images. Based on the fact that sharp edges in the image correspond to concentrated gradients along the edge, Fattal *et al*. [7] model the edge profile distribution. Sun *et al*. [19] exploit the gradient transformation from HR image to LR image. And Tai *et al*. [20] recover the fine gradient using the user-supplied exemplar texture. Nevertheless, it is hard to model the image with a few parameters. Therefore the reconstructed images are usually over-sharped or suffer from false artifacts due to the incorrect gradient estimation. Our method follows a similar way to recover the image gradient, and we exploit the useful information from the external gradient patterns, thus being more expressive than the parameter-controlled gradient prior methods.

Patch based methods became popular for the simplicity to represent the local structure. Patches or epitomes have been successfully used in image labeling and classification [3, 16]. Patches' self-similar properties are exploited in the work [8, 11, 14, 29]. But in this paper we focus on the external example based methods, in which a universal set is used to provide numerous texture patterns for HR details prediction. As in the seminal work of Freeman *et al*. [9], Markov Random Field is employed to select the appropriate HR patch from a bundle of candidates. Chan *et al*. [2] propose a Neighbor Embedding method inspired by LLE algorithm in manifold learning, followed by extensions of this work [10, 28]. Sparse coding methods [24, 25] exploit the sparsity property in the patch representation. The basic assumption is that HR/LR patches share the coefficients when they are sparsely coded by HR/LR dictionary. The dictionary is usually trained to be compact to represent the HR patterns. He *et al*. [13] extend the work by allowing a mapping function between HR and LR sparse coefficients. Timeofte *et al*. [21] also propose an improved variant of Anchored Neighborhood Regression (ANR) method. Nevertheless, these methods use patches as a fixed vector. This requires an extremely large dictionary to cover the input patch structure or linear combination components. To relax the problem, Zhu *et al*. [30] allow a patch deformation, thus making the dictionary more expressive.

Recently, deep learning showed its power to learn hierarchical representations of high dimensional data, and it is already being used in example based super-resolution. Cui *et al*. [4] follow the gradual upscaling strategy from [8, 11] and apply the Collaborative Local Auto-encoder (CLA) to introduce the information of the external data in each step. Thus, they build a deep network cascade reconstruction structure. And Dong *et al*. [6] extend the sparse coding method to a convolutional neural network(CNN) learning structure with different mapping layers. But these methods seldom impose any gradient constraint on the desired image. So there is still a blurring effect that could be improved. Moreover, none of these method exploits what kind of structures is useful for super-resolution. Instead, they put all the patches into the learning or reconstruction process.

There are also related work on image enhancement [12, 17, 18]. These methods tend to sharpen the discontinuous edges independently. Thus the results are not so reasonable to human perception. The drawback is mentioned in the literature [19].

In our work, we applied deformable patch based method [30] to make the gradient pattern dictionary more expressive. By the decomposition into singular primitive structures, the deformable patch method is more robust, which is a big improvement over the deformable patch based method, as shown in our experiments.

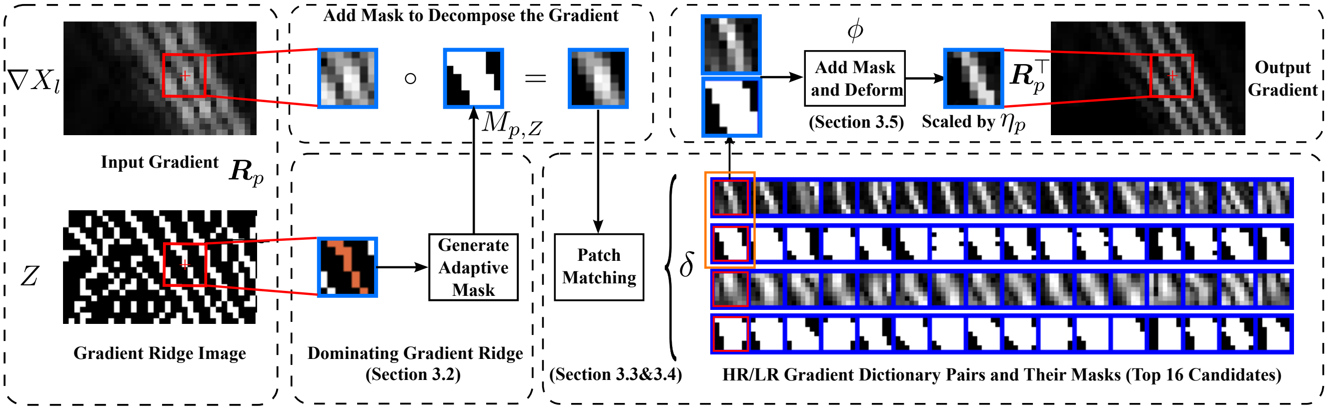


Figure 2. Flow chart of the HR gradient field estimation. The input gradient ridge image is obtained from the input gradient field. For a pixel at position p (marked with a red cross), we first generate the mask from the dominating gradient ridge (orange part). Then the LR gradient is integrated with the mask to extract the singular part. We find the best candidate in the dictionary (marked by red box) and the corresponding masks. Finally the masked HR gradient patches are deformed for the final estimation. Note that here we simply show the gradient magnitude for illustration, but actually the x and y components are calculated separately and simultaneously, for both the input images and the dictionary elements. You may see the details in the corresponding section.

3. Deformable Gradient Compositional Model

Previous example based methods do not work well on the non-singular patterns (See Figure 1). In this section, we introduce the Gradient Compositional Model, aiming to focus on each singular part of the non-singular patterns by adding masks. We also allow the deformation for the HR patches to make them fit the LR input flexibly. The flow chart is shown in Figure 2. In Section 3.1 we overview our proposed model and we will discuss the details in the subsequent subsections.

3.1. Gradient Compositional Model

To decompose the patches into non-singular structures, 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. This is equivalent to the zero crossing map in the work [19]. We will elaborate on how to get gradient ridges in Section 3.2.

First of all, 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 $D_{gl} = (D_{xl}, D_{yl})$ (the raw gradient dictionary elements and the masks are displayed separately in Figure 2), 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. As a summary, suppose we have an input LR image X_l and its gradient ∇X_l. The predicted HR gradient $\nabla\tilde{X}_h$ is reconstructed as the following equation:

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

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.

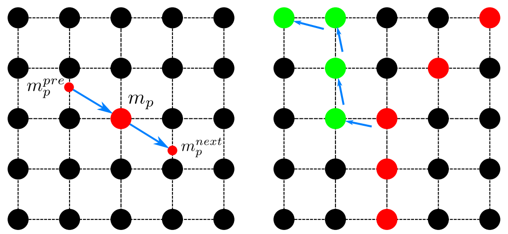
$M_{p,Z}$: the patch mask generated from the gradient ridge point, centered at p;

$\boldsymbol{\delta}$: the indicative function that chooses the best match between the input LR patch and LR gradient dictionary (See Section 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 the corresponding gradient dictionary is also used in edge smoothness estimation (Section 3.3) and patch matching (Section 3.4). More details can be found in the experiments Section 5.1. 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 2.



(a) Find the gradient ridges (b) Find the mask

Figure 3. (a) Gradient ridges searching strategy, for arbitrary pixels p (big red dot), the previous and next positions (small red dots) are found along and against the gradient direction (blue arrow). (b) Mask marking strategy. The red dots denote the gradient ridge pixels. The green dots indicate the searching path from the corner pixels. Blue arrows are gradient directions of each green point. In this case, starting from the top left, the path finally reaches a red dot. So they should be marked as mask pixels.

3.2. Gradient Ridges and Mask Generation

In this section, we describe how to obtain the gradient ridge pixels Z from an input image X. Suppose we have a gradient field in the form $\nabla X = m \cdot \overrightarrow{N}$, where m is the gradient magnitude and $\overrightarrow{N}$ is the gradient direction. Along the gradient direction, we can find the next sub-pixel position with the gradient magnitude m_p^{next} determined by linear interpolation of the neighbors. Similarly, against the gradient direction, we find the previous position and gradient magnitude m_p^{pre}. If m_p is the maximum among them, it is marked as a gradient ridge pixel (Figure 3(a)). The same strategy is used in [19] for finding the edge pixel. The following equation shows how to get the gradient ridges Z_p:

$$Z_p = \begin{cases} 1 & m_p > m_p^{pre} \;\&\; m_p > m_p^{next} \\ 0 & otherwise \end{cases} \tag{2}$$

Given a patch with the center at a gradient ridge pixel p, first we find the connected ridge pixels that cover the center pixel by 8-connected searching, *i.e.* the dominating ridge (mark in orange color in Figure 2). Next, we search for the mask to delimit the dominating area of the dominating ridge. The dominating area starts from the gradient ridge and ends at the gradient magnitude local minimum along the edge direction. Therefore we mark all the dominating ridge pixels as mask pixels. And for each pixel within the patch, we search for the path along the increasing gradient magnitude direction, *i.e.* against the gradient direction $(\partial m/\partial x, \partial m/\partial y)$. Here we just choose the next pixel by the number of $\partial m/\partial x$ and $\partial m/\partial y$ to make the path go through every pixel. That is to say, if $|\partial m/\partial x| > |\partial m/\partial y|$, there should be $x^{next} = x + sign(\partial m/\partial x)$ for a fixed y. If the path reaches a dominating ridge pixel, all the pixels on the path are marked as mask pixels. Otherwise they will not be marked if the path reaches the border or loops on a single pixel. See Figure 3(b).

In this paper, we use two kinds of patches, the gradient patches and the intensity patches. When integrated to the gradient patches, the gradient pixels covered by the mask are kept and the rest are set to 0. But when we add the mask to the intensity patches, we adopt a different strategy. For the unmasked pixels within the intensity patch, we fill them from the boundary of the masks iteratively. Then for each unmasked pixel, if 3 masked pixels are found with it 8-connected area, the unmasked pixel is assigned the average of these pixels and set as a masked pixels. This processing is repeated until all the pixels are masked.

3.3. Edge Smoothness

Gradient ridges describe the positions and shapes of the singular structures. But they do not keep the edge smoothness information, which is very important for patch matching (Section 3.4). It helps to find the HR patches with appropriate edge width. Obviously, we can not expect all reconstructed edges to be as sharp as possible. Instead, it should be adaptive to the input LR gradient.

We estimate the edge smoothness as weighted combined distance. The weight is defined as the normalized gradient magnitude m_p within the mask area. Thus the edge smoothness has the form of the variance of the edge profile (as in [19]). Within a patch (HR or LR), for the pixel located at position p, it corresponds to a ridge pixel p_z on the searching path, accordingly with Section 3.2. The smoothness S is formulated by the weighted combination of the distance between p and p_z:

$$S = \frac{\sum_{p \in M} (p - p_z)^2 m_p}{\sum_{p \in M} m_p} \tag{3}$$

where M denotes the masks within the patch, m_p is the gradient magnitude on pixel p. Since we integrate on all the masked pixels, the smoothness is a scalar for each patch.

3.4. Patch Matching

Patch matching is an essential stage in example based super-resolution methods when HR/LR paired dictionary is employed. In the literature [9], MRF is utilized for the best candidate selection. Also the linear combination methods (*e.g.* sparse coding [13, 21, 24], neighbor embedding [2, 28] *et al.*) and deformable patch based methods [30] involve the patch selection or matching step. In this paper, we match the input LR gradient patch with the LR part of the dictionary, aiming to find the indicative function $\boldsymbol{\delta}$, which indicates the single HR patch that are suitable for deformation instead of the exact version. We consider two principles:

1. The patch with more similar edge orientation is more likely to be selected. The orientation can be represented by the histogram of gradient (HOG) [5] of the masked intensity patches.

2. Edge smoothness in Section 3.3 is considered to select the dictionary element with similar LR edge width to the input LR patches:

Based on the above, given the input masked LR gradient patch $P_{gl} = M_{p,Z}\boldsymbol{R}_p \nabla X_l$, first we find its corresponding masked intensity patch $P_l = M_{p,Z}\boldsymbol{R}_p X_l$. Meanwhile LR intensity dictionary D_l is corresponded to the gradient dictionary D_{gl}. Thus the indicative function $\boldsymbol{\delta}$ is defined as:

$$\arg\min_{\boldsymbol{\delta}} ||Hog(P_l) - Hog(D_l)\boldsymbol{\delta}||^2 + \lambda|S(P_{gl}) - S(D_{gl})\boldsymbol{\delta}|^2 \tag{4}$$

where $Hog(\cdot)$ is the HOG feature descriptor. $S(\cdot)$ is the patch smoothness described in Section 3.3. λ is the balance parameter. In our experiments, we set $\lambda = 1$.

3.5. Patch Deformation

Recent work of deformable patches [30] shows that by allowing the patch deformation, the dictionary can be made more flexible and expressive. They assume that during the degradation, the HR/LR pairs share the deformation field instead of the sparse linear combination coefficients as [13, 25] do.

In this paper, we apply the deformable patches on the intensity patches rather than the gradient patches. The reason is that the basis of the deformation in [30] is a Taylor expansion which is satisfied when the patch is smooth and continuous. For the gradient patches case, they do not maintain such a good continuous property, which may lead to large errors. Therefore the deformation field estimation stage does not work well with the discrete gradient patches.

In the approach we propose, first we add the mask to the intensity patches to make them contain a single edge as described in Section 3.2. The dictionary elements are masked in the same way. After adding the masks, both the input LR patch and the selected HR candidate contain singular structures. Then we minimize the following energy function to obtain the deformation field u (along x direction) and v (along y direction):

$$E = ||B_l + B_{lx} \circ u + B_{ly} \circ v - P_l||^2 + \psi(u, v) \tag{5}$$

Under the assumption that the LR/HR patches share the deformation field, the warping function ϕ has the form:

$$\phi(B_h) = \nabla(B_h + B_{hx} \circ u + B_{hy} \circ v), \tag{6}$$

where P_l is the masked input intensity patch. $B_h = D_h\boldsymbol{\delta}$ is the selected HR candidate for input LR patch. Similarly $B_l = D_l\boldsymbol{\delta}$. $\boldsymbol{\delta}$ is defined in Section 3.4. B_{lx} and B_{ly} are the gradients along x and y respectively, B_{hx} and B_{hy} is the same for HR patches. Operator $\circ$ denotes the point-wise multiplication. We also use $\psi(u, v)$ as the second-order derivative constraint as [30] does. Note that the final ϕ is normalized in Eqn. (1). So η_p is introduced to adjust the contrast.

4. Super-resolution Reconstruction Model

In the previous section we presented the deformable gradient compositional model. By focusing on the single edge in each patch covered by its mask, we can obtain more reliable gradient estimation. These gradient patches are integrated on $\nabla\tilde{X}_h$ as in Eq.(1) ($\nabla_x\tilde{X}_h$ and $\nabla_y\tilde{X}_h$ are processed separately). In this section, 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 [19, 20] The following energy function is minimized by enforcing the constraint in both intensity domain and gradient domain:

$$E(X_h|X_l, \nabla\tilde{X}_h) = ||SHX_h - X_l||^2 + \beta||\nabla X_h - \nabla\tilde{X}_h||^2, \tag{7}$$

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. In this paper, we choose $\beta = 0.3$. The above energy function is quadratic and convex, so the global minimum can be obtained by gradient decent:

$$\begin{aligned} X_h^{t+1} =& X_h^t - \tau[(H^\top S^\top SHX_h - H^\top S^\top X_l) \\ & - \beta(\mathrm{div}(\nabla X_h) - \mathrm{div}(\nabla\tilde{X}_h))] \end{aligned} \tag{8}$$

where t is the iteration number and τ is the iteration step, which we set to $\tau = 0.1$. We have tried different value ranging in $[0, 1]$ for β and τ, but the results remain stable. $\mathrm{div}(\nabla X_h)$ denotes the divergence of ∇X_h via the form $\mathrm{div}(\nabla X_h) = \partial^2 X_h/\partial x^2 + \partial^2 X_h/\partial y^2$, which can be implemented easily using the Laplace operator. In this paper, we choose CNN super-resolution [6] result for the X_0 initialization.

5. Experimental Results

In this section, we evaluate our algorithm on the test examples commonly used in the super-resolution literature. We evaluate the deformable gradient compositional model and discuss its contribution over the deformable patches.

5.1. Dictionary and Experiments Setting

Dictionary Preparation In our experiment, the dictionary is generated from the Berkeley Segmentation Dataset 500 [12] by randomly selecting enormous patch pairs as in [9, 13, 23, 25, 30].

For each HR image I_h in the dataset, we calculate the LR version $I_l = H^\top S^\top SHI_h$. Here $H^\top S^\top$ is an upscaling process, *i.e.* we use the upscaled version of the low-resolution image. Thus, the patch size of the LR and HR patches can be fixed as the same. The literature [23, 25, 30] adopts the same strategy. $H^\top S^\top$ also appears when iteratively updating X_h in Eq. (8). In this paper, we use bicubic interpolation to implement the upscaling.

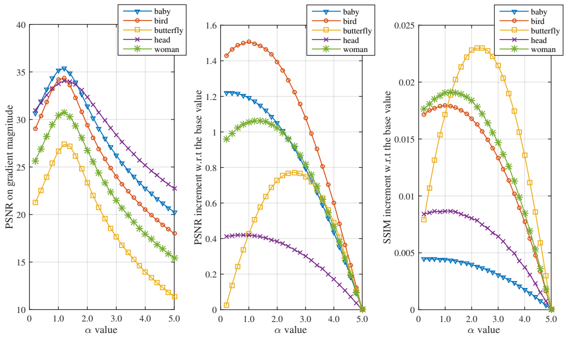


Figure 4. PSNR and SSIM for different α. Left: PSNR on gradient magnitude. Middle: Final PSNR increment w.r.t. the base value baby (34.07dB), bird (33.78dB), butterfly (27.23dB), head (33.22dB), woman (30.11dB). Right: Final SSIM increment w.r.t the base value baby (0.9711), bird (0.9348), butterfly (0.8899), head (0.8169), woman (0.9073).

Since our deformable gradient compositional model involves the gradient patches, the corresponding gradient images $\nabla I_h = (\nabla_x I_h, \nabla_y I_h)$, $\nabla I_l = (\nabla_x I_l, \nabla_y I_l)$ are generated. Then, we randomly select the LR patches that are centered at a LR gradient ridge pixels. From the same position, all the 2 intensity patches and 4 gradient patches are extracted as one set $\{D_h, D_l, D_{x_h}, D_{x_l}, D_{y_h}, D_{y_l}\}$. All of the dictionary patches are integrated with their corresponding masks.

Parameter Setting In the experiments, the patch size is 7×7 and the upscale factor is $3\times$. Initially, we randomly select 30000 patch sets (including intensity patches and corresponding gradient patches). In the reconstruction step, we set $\beta = 0.3$ and $\tau = 0.1$. Finally, we incorporate the non-local method [1] for post processing as work [13, 30] do. For color images, super-resolution is done on Y channel in the YCbCr color space, and the other two channels are upscaled by bicubic interpolation. In Section 5.2 and 5.3 we evaluate the performance in term of PSNR and SSIM [22], High PSNR/SSIM indicate good performance. For each image, we down-sample the ground truth image first and upscale it by super-resolution methods. PSNR and SSIM are calculated between the result and the ground truth. 5-pixel borders are excluded because we do not process the border pixels.

Testing Images 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 [6, 21, 26]. In the parameter chosen and effectiveness section, five images that compose Set 5 are used. And in the final super-resolution evaluation section, all the 19 images from Set 5 and 14 are involved.

5.2. Contrast adjustment parameter α

In this section, we exploit how to choose the global contrast adjustment parameter α in Section 3.1. The final performance is sensitive to this parameter. Here we test the possible choices between 0.2 and 5.0 with a 0.2 step. These 5 test images are from Set 5. The other parameter is set as in the previous section. PSNR on the gradient magnitude is displayed in Figure 4 (Left), which represents the gradient reconstruction error. Figure 4 (Middle and Right) show PSNR/SSIM of the final results in the form of increment with regard to the base value because of the different range for different images. The base values are the minimums of the 25 points on each curve. Different curves represent the performance on each image. Thus they are not comparable. The trend is only observed on a single curve.

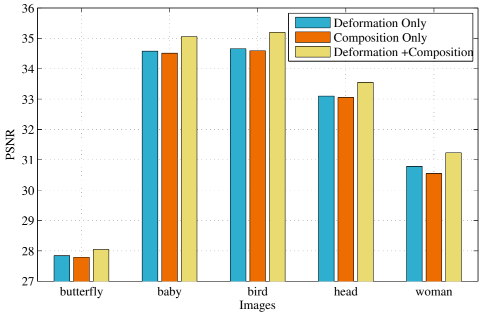


Figure 5. The effectiveness evaluation of the patch deformation and the gradient composition.

In Figure 4 (Left), a peak at around $\alpha = 1.2$ can be observed. Meanwhile at the final PSNR/SSIM curves (Figure 4 Middle and Right) show similar trends at $\alpha = 1.2$. For each curve, when α is small (< 1.2), PSNR/SSIM decrease as α decreases, but they remain higher than the case of big α (> 4). This is reasonable. For a small α, the gradient constraint suppress the magnitude of the gradient, which is equivalent to the smoothness prior. Meanwhile, Eqn. (7) can be regarded as a back-projection form with smoothness prior. This process can also help improve PSNR performance, because the error term guarantees the fidelity to the input LR image and the smoothness reduces the possible error due to misaligned edges. However, in these cases the edges are usually blurred and the gradient constraint is not imposed. In our observation, butterfly image is an exception due to its different gradient statistics (*i.e.* more edges than others). But at $\alpha = 1.2$ it still has good PSNR/SSIM. Therefore we choose $\alpha = 1.2$ as the universal setting.

5.3. Effectiveness of the model

We conduct experiments to evaluate the effectiveness of the proposed deformable compositional model. The 5 images in Set 5 are used for evaluation. First, we ignore the gradient composition *i.e.* set all of the masks to be all-1 patch for both the input patch and the dictionary. The patch matching also involves only the unmasked patches. And next we make the code go through the compositional model, but without deformation. Instead, the top candidate during patch matching is directly used. The proposed Deformable Compositional Model results are added for com-

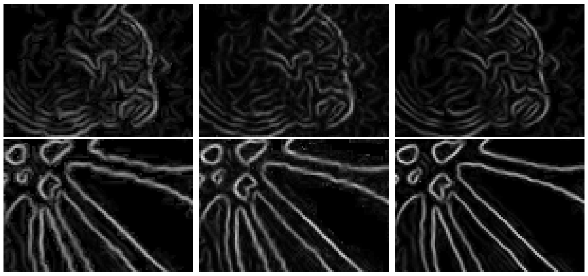


Figure 6. The gradient map generated by deformation only, composition only and the proposed decompositions model (from left to right). The top images are from the head image in Set 5 and the bottom ones are from the butterfly image.

parison. Figure 5 displays the PSNR performance on the 3 cases. From the figure, we can see that the deformation makes a little more contribution than the compositions, but not much. Both of them contribute to the final performance.

Figure 6 displays two example parts from the test images. For the deformation only and composition only cases, the reconstructed gradient suffers from the blurring around the edges. In the final deformable gradient compositional model result, the edge is more concentrated, avoiding the blur and sticking edges.

5.4. Super-resolution results

To evaluate the performance, in this section, we choose 19 test examples (Set 5 and Set 14) used in the previous literature [6, 21, 26]. We will show the PSNR/SSIM quantitative results and visual quality. More specifically, Four categories of methods are selected as the competitors, including Gradient Profile Prior (GPP [19]), dictionary learning methods (*i.e.* Sparse Coding [25], and Anchored Neighborhood Regression ANR [21]), Deformable Patches (DPSR [30]) as well as the deep networks(*i.e.* Convolutional Neural Network CNN [6] and Deep Network Cascade DNC [4]). Table 1 shows the PSNR and SSIM results of the competitors and our method. Overall, our method outperforms the state-of-the-art methods. We outperform the other methods on 14 of total 19 images in PSNR. And we achieve higher SSIM in 16 of all the 19 cases due to more appropriate gradient constraint. Figure 8 displays some parts of the images and the corresponding gradient magnitude image. This part is chosen to show the super-resolution results on non-singular structure patterns. Visually, in the part of the butterfly, both the dictionary learning based method [21, 25] and the deformable patch method [30] suffer seriously from the bicubic jag artifact, due to the linear or weighted combination of the dictionary elements. The Deep learning method [6] shows smoother results than the others due to the non-linear mapping between the layers, but it fails to maintain sharp edges around the non-singular areas since they do not consider the gradient constraints as [19] and our method does. On the contrary, [19] can generate sharp edges, but the gradient field is not natural because it is hard to model the complex edge composition using a few parameters. Figure 7 shows more results on three images. Four recent methods are compared. Overall, our method can generate better natural gradient field and sharp images with the help of the gradient deformable compositional model.

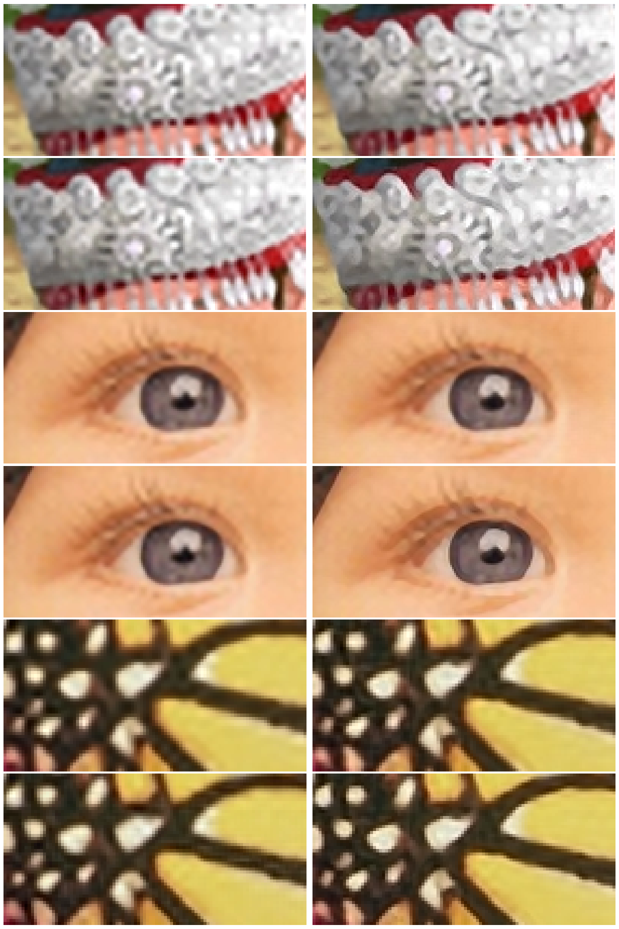


Figure 7. We show more comparisons with the state-of-the-art methods, 3×. For each group, from left to right, from top to bottom: DPSR [30], DNC [4], CNN [6], and our method. Our method performs better with sharp edges. Please zoom in for better view.

6. Conclusion

In this paper, we propose a Deformable Gradient Compositional model to represent the non-singular primitives as compositions of singular ones, each of which is allowed some deformation. The compositional aspect makes the gradient field more reliable. The deformation aspect makes the dictionary more expressive. The experiments on Set 5 and Set 14 show that our method outperforms the state-of-the-art methods. 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.

Table 1. Performance in PSNR and SSIM on the 19 images in Set 5 and Set 14. Upscale factor: 3×

Images	GPP[19]		SCSR[25]		ANR[21]		DPSR[30]		SRCNN[6]		DNC[4]		Proposed	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
baboon	23.41	0.8470	23.57	0.8529	23.56	0.8514	23.58	0.8533	23.59	0.8528	23.58	0.8546	**23.63**	**0.8565**
baby	34.38	0.9716	35.15	0.9748	35.07	0.9745	35.17	0.9751	34.88	0.9733	34.92	0.9749	**35.24**	**0.9755**
barbara	26.59	0.8970	26.80	**0.9007**	26.65	0.8974	**26.80**	0.9005	26.60	0.8962	26.60	0.8975	26.73	0.8995
bird	33.56	0.9335	34.83	0.9492	34.63	0.9479	34.86	0.9496	34.89	0.9466	34.99	0.9496	**35.28**	**0.9527**
bridge	24.83	0.8987	25.09	0.9044	25.02	0.9021	25.11	0.9049	25.07	0.9027	25.18	0.9043	**25.20**	**0.9068**
butterfly	26.29	0.8820	26.15	0.8789	25.92	0.8726	26.60	0.8878	27.61	0.9007	27.08	0.9009	**27.77**	**0.9088**
coastguard	27.04	0.6553	**27.25**	**0.6645**	27.14	0.6561	27.21	0.6640	27.17	0.6559	27.14	0.6608	27.23	0.6643
comic	23.85	0.7562	24.04	0.7649	23.98	0.7587	24.04	0.7646	24.31	0.7749	24.20	0.7674	**24.41**	**0.7814**
face	33.06	0.8120	33.53	0.8230	33.55	0.8219	33.58	0.8243	33.48	0.8185	33.57	0.8238	**33.66**	**0.8256**
flowers	28.21	0.8327	28.53	0.8429	28.42	0.8390	28.61	0.8444	28.88	0.8445	28.87	0.8488	**29.06**	**0.8525**
foreman	33.45	0.9269	33.75	0.9320	33.59	0.9309	33.98	0.9337	33.95	0.9309	34.42	0.9370	**34.42**	**0.9376**
head	33.16	0.8141	33.58	0.8244	33.57	0.8227	33.61	0.8251	33.44	0.8181	**33.65**	0.8252	33.64	**0.8256**
lena	32.45	0.9639	33.09	0.9674	33.02	0.9672	33.17	0.9678	33.30	0.9676	33.28	0.9684	**33.48**	**0.9694**
man	27.73	0.9327	27.97	0.9367	27.89	0.9353	28.00	0.9368	28.12	0.9367	28.20	0.9395	**28.25**	**0.9399**
monarch	31.10	0.9791	31.22	0.9809	31.03	0.9806	31.57	0.9816	32.28	0.9825	31.93	0.9825	**32.47**	**0.9837**
pepper	33.56	0.9752	34.10	0.9778	33.83	0.9770	34.12	0.9777	34.27	0.9771	34.50	0.9784	**34.51**	**0.9788**
ppt3	24.71	0.9616	25.09	0.9641	24.87	0.9614	25.14	0.9644	25.80	0.9673	**26.15**	**0.9732**	25.83	0.9694
woman	29.83	0.9061	30.52	0.9187	30.36	0.9161	30.50	0.9183	30.89	0.9201	**31.23**	0.9261	31.17	**0.9264**
zebra	28.05	0.8331	28.68	0.8500	28.38	0.8427	28.69	0.8508	28.77	0.8456	29.03	0.8494	**29.28**	**0.8554**
Average	29.22	0.8831	29.63	0.8899	29.50	0.8871	29.70	0.8908	29.86	0.8901	29.92	0.8927	**30.07**	**0.8952**

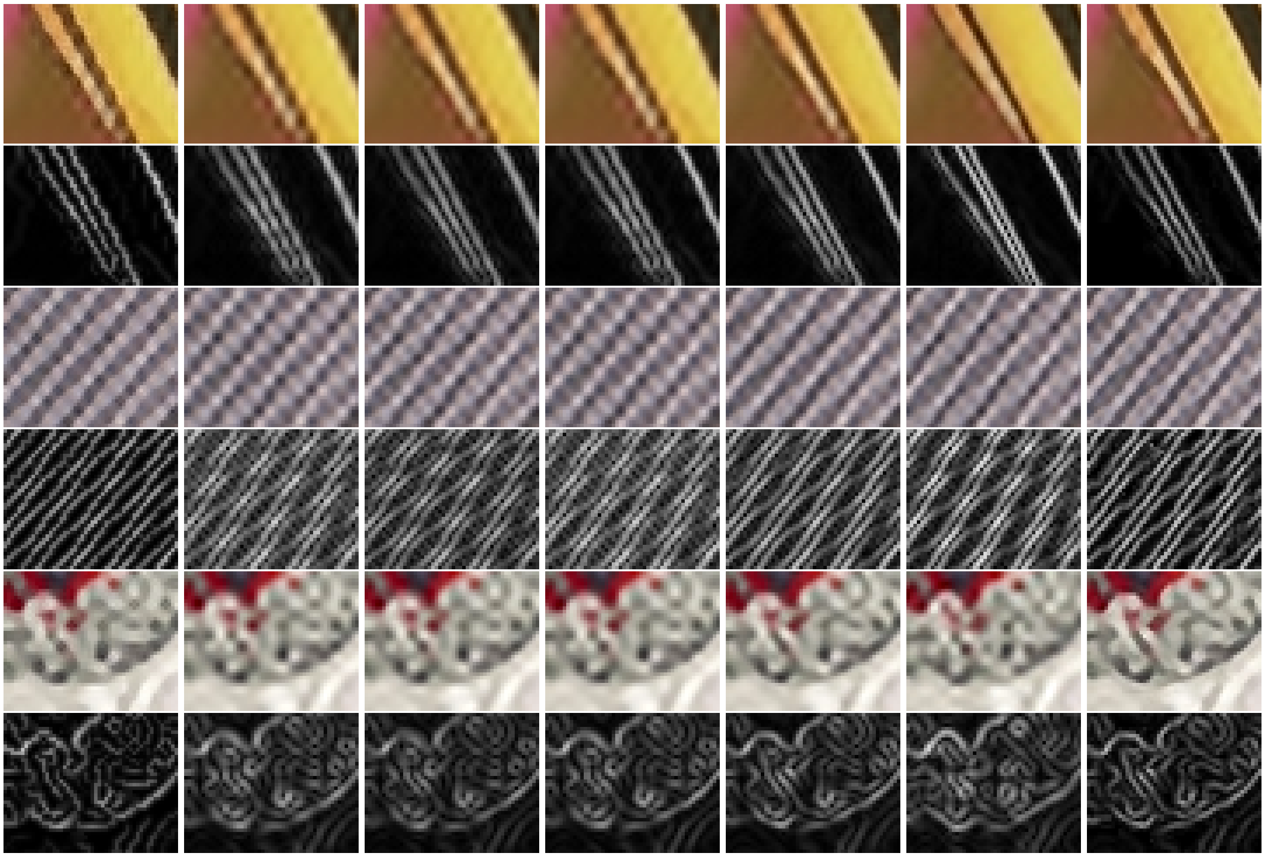


Figure 8. Results on some non-singular parts in Set 5 and Set 14, the corresponding gradient field is displayed. Upscale factor: 3×. Image from top to bottom: butterfly, barbara and comic. Please zoom in for better view.

Acknowledgement 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.

References

[1] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *CVPR*, volume 2, pages 60–65, June 2005. 1, 6

[2] T.-M. Chan, J. Zhang, J. Pu, and H. Huang. Neighbor embedding based super-resolution algorithm through edge detection and feature selection. *PR Letters*, 30(5):494–502, 2009. 1, 2, 4

[3] L.-C. Chen, G. Papandreou, and A. Yuille. Learning a dictionary of shape epitomes with applications to image labeling. In *ICCV*, pages 337–344, Dec 2013. 2

[4] Z. Cui, H. Chang, S. Shan, B. Zhong, and X. Chen. Deep network cascade for image super-resolution. In *ECCV*, 2014. 1, 2, 7, 8

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893 vol. 1, June 2005. 4

[6] C. Dong, C. C. Loy, K. He, and X. Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014. 1, 2, 5, 6, 7, 8

[7] R. Fattal. Image upsampling via imposed edge statistics. In *SIGGRAPH*, SIGGRAPH '07, 2007. 1, 2

[8] G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Trans. Graph.*, 30(2):12:1–12:11, 2011. 1, 2

[9] W. Freeman, T. Jones, and E. Pasztor. Example-based super-resolution. *Computer Graphics and Applications*, 22(2), 2002. 1, 2, 4, 5

[10] X. Gao, K. Zhang, D. Tao, and X. Li. Image super-resolution with sparse neighbor embedding. *IEEE TIP*, 21(7):3194–3205, 2012. 2

[11] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *ICCV*, pages 349–356, 2009. 1, 2

[12] F. Guichard and J.-M. Morel. A note on two classical enhancement filters and their associated pde's. *IJCV*, 52(2-3):153–160, 2003. 2, 5

[13] L. He, H. Qi, and R. Zaretzki. Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution. In *CVPR*, pages 345–352, 2013. 1, 2, 4, 5, 6

[14] J. Hu and Y. Luo. Single-image superresolution based on local regression and nonlocal self-similarity. *Journal of Electronic Imaging*, 23(3):033014, 2014. 2

[15] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE TPAMI*, 32(6):1127–1133, June 2010. 1, 2

[16] G. Papandreou, L.-C. Chen, and A. Yuille. Modeling image patches with a generic dictionary of mini-epitomes. In *CVPR*, pages 2059–2066, June 2014. 2

[17] A. Polesel, G. Ramponi, and V. Mathews. Image enhancement via adaptive unsharp masking. *IEEE TIP*, 9(3):505–510, Mar 2000. 2

[18] L. A. Rudin. *Images, Numerical Analysis of Singularities and Shock Filters*. PhD thesis, California Institute of Technology, Pasadena, CA, USA, 1987. UMI Order No. GAX87-19699. 2

[19] J. Sun, J. Sun, Z. Xu, and H.-Y. Shum. Gradient profile prior and its applications in image super-resolution and enhancement. *TIP*, 20(6):1529–1542, June 2011. 1, 2, 3, 4, 5, 7, 8

[20] Y.-W. Tai, S. Liu, M. S. Brown, and S. Lin. Super resolution using edge prior and single image detail synthesis. In *CVPR*, pages 2400–2407, 2010. 2, 5

[21] R. Timofte, V. De, and L. Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *ICCV*, pages 1920–1927, Dec 2013. 1, 2, 4, 6, 7, 8

[22] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, April 2004. 6

[23] J. Yang, Z. Lin, and S. Cohen. Fast image super-resolution based on in-place example regression. In *CVPR*, pages 1059–1066, 2013. 1, 2, 5

[24] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *CVPR*, pages 1–8, 2008. 1, 2, 4

[25] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE TIP*, 19(11):2861–2873, 2010. 1, 2, 5, 7, 8

[26] 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. 1, 6, 7

[27] H. Zhang, J. Yang, Y. Zhang, and T. S. Huang. Non-local kernel regression for image and video restoration. In *ECCV*, pages 566–579, 2010. 2

[28] K. Zhang, X. Gao, X. Li, and D. Tao. Partially supervised neighbor embedding for example-based image super-resolution. *IEEE Journal of Selected Topics in Signal Processing*, 5(2):230–239, 2011. 1, 2, 4

[29] K. Zhang, X. Gao, D. Tao, and X. Li. Single image super-resolution with non-local means and steering kernel regression. *IEEE TIP*, 21(11):4544–4556, Nov 2012. 2

[30] Y. Zhu, Y. Zhang, and A. L. Yuille. Single image super-resolution using deformable patches. In *CVPR*, pages 2917–2924, 2014. 1, 2, 4, 5, 6, 7, 8