

Visualizing Image Priors

Tamar Rott Shaham^(✉) and Tomer Michaeli

Technion—Israel Institute of Technology, Haifa, Israel
stamarot@campus.technion.ac.il, tomer.m@ee.technion.ac.il

Abstract. Image priors play a key role in low-level vision tasks. Over the years, many priors have been proposed, based on a wide variety of principles. While different priors capture different geometric properties, there is currently no unified approach to interpreting and comparing priors of different nature. This limits our ability to analyze failures or successes of image models in specific settings, and to identify potential improvements. In this paper, we introduce a simple technique for visualizing image priors. Our method determines how images should be deformed so as to best conform to a given image model. The deformed images constructed this way, highlight the elementary geometric structures to which the prior resonates. We use our approach to study various popular image models, and reveal interesting behaviors, which were not noticed in the past. We confirm our findings through denoising experiments. These validate that the structures we reveal as ‘optimal’ for a specific prior are indeed better denoised by this prior.

1 Introduction

Image priors play a fundamental role in many low-level vision tasks, such as denoising, deblurring, super-resolution, inpainting, and more [1–9]. Over the years, many priors have been proposed, based on a wide variety of different principles. These range from priors on derivatives [2, 10], wavelet coefficients [11, 12], filter responses [13, 14], and small patches [1, 15], to nonparametric models that rely on the tendency of patches to recur within and across scales in natural images [16–19].

Different priors capture different geometric properties. For example, it is known that the total variation (TV) regularizer [10] prefers boundaries with limited curvature [20], whereas the local self-similarity prior [21] prefers straight edges and sharp corners (structures which look the same at different scales). However, generally, characterizing the behavior of complex image priors (*e.g.*, trained models) is extremely challenging. This limits our ability to interpret failures or successes in specific settings, as well as to identify possible model improvements.

In this paper, we present a simple technique for visualizing image priors. Given an image model, our method determines how images should be deformed so that they become more plausible under this model. That is, for any input image, our algorithm produces a geometrically ‘idealized’ version, which better

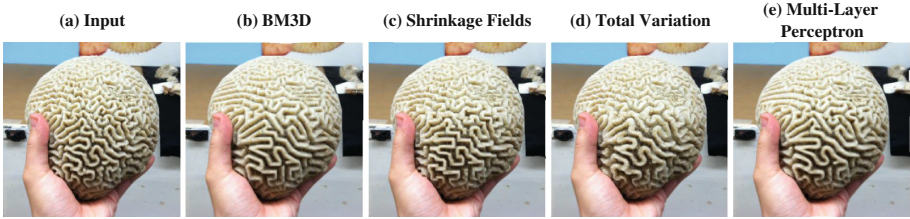


Fig. 1. Visualizing image priors. Our algorithm determines how images should be deformed so as to better comply with a given image model (exemplified here on a Brain Coral image). The deformed images give insight into the elementary geometric features to which the prior resonates. As can be seen, different image models (BM3D [17], Shrinkage Fields [22] with pairwise cliques, Total Variation [10], Multi-Layer Perceptron [4]) have quite different geometric preferences.

conforms to the prior we wish to study. Figure 1 shows several example outputs of our algorithm. As can be seen, our idealization process nicely highlights the elementary features to which different priors resonate, and thus gives intuition into their geometric preferences.

Our approach is rather general and, in particular, can be used to visualize generative models (*e.g.*, fields of experts [14]), discriminative models (*e.g.*, deep nets [4]), nonparametric models (*e.g.*, nonlocal means [16]), and any other image model that has an associated denoising algorithm. In fact, the ‘idealized’ images produced by our method have a nice interpretation in terms of the associated denoiser: Their geometry is not altered if we attempt to ‘denoise’ them (treating them as noisy images). We thus refer to our ‘idealized’ images as *Geometric Eigen-Modes* (GEMs) of the prior.

Figure 2 illustrates how GEMs encode geometric preferences of image models. For example, since the TV prior [10] penalizes for large gradients, a TV-GEM is a deformed image in which the gradient magnitudes are smaller. Similarly, the wavelet sparsity prior [11] penalizes for non-zero wavelet coefficients. Therefore, a wavelet-GEM is a deformed image in which the wavelet coefficients are sparser. Finally, the internal KSVD model [15] assumes the existence of a dictionary over which all patches in the image admit a sparse representation. Thus, a KSVD-GEM is a deformed image for which there exists a dictionary allowing better sparse representation of the image patches.

We use our approach to study several popular image models and observe various interesting phenomena, which, to the best of our knowledge, were not pointed out in the past. First, unsurprisingly, we find that all modern image priors prefer large structures over small ones. However, the preferred shapes of these large objects, differ among priors. Specifically, most internal priors (*e.g.*, BM3D [17], internally-trained KSVD [15], cross-scale patch recurrence [19]) prefer straight edges and sharp corners. On the other hand, externally trained models (*e.g.*, EPLL [1], multi-layer perceptron [4]), are much less biased towards straight borders, and their preferred shapes of corners are rather round. But we also find a

few surprising exceptions to this rule. For example, it turns out that nonlocal means (NLM) [16], which is an internal model, rather resonates to curved edges, similarly to external priors. Another interesting exception is the fields of experts (FoE) prior [14], an externally-trained model which turns out to prefer straight *axis-aligned* edges.

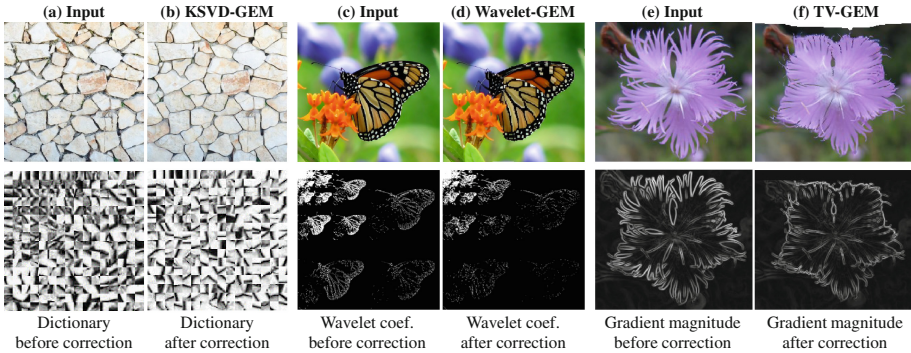


Fig. 2. GEMs better conform to the prior. (a) The internal KSVD model [15] assumes that each patch in the image can be sparsely represented over some dictionary. (b) A KSVD-GEM is a deformed image in which the diversity between patches is smaller, so that the sparsity assumption holds better. Namely, for the KSVD-GEM, there exists a dictionary over which each patch can be sparsely represented *with better accuracy*. Note how less atoms are invested in representing the fine details in this dictionary. (c) The wavelet sparsity prior [11] penalizes the ℓ_1 norm of the wavelet coefficients of the image (we use the Haar wavelet for illustration). (d) A wavelet-GEM is a deformed image in which the wavelet coefficients have a smaller ℓ_1 norm, and are thus sparser. (e) The TV prior penalizes the ℓ_1 norm of the gradient magnitude. (f) A TV-GEM is a deformed image in which the gradient magnitude is smaller (and so has a smaller ℓ_1 norm).

The behaviors we reveal are often impossible to notice visually in standard image recovery experiments on natural images (*e.g.*, denoising, deblurring, super-resolution). However, they turn out to have significant effects on the PSNR in such tasks. We demonstrate this through several denoising experiments. As we show, structures predicted by our approach to be most ‘plausible’, can indeed be recovered from their noisy versions significantly better than other geometric features. So, for example, we show how the FoE model indeed performs significantly better in denoising an axis-aligned square, than in denoising a rotated one.

1.1 Related Work

There are various approaches to interpreting and visualizing image models. However, most methods are suited only to specific families of priors, and are thus of limited use when it comes to *comparing* between models of different nature.

Moreover, existing visualizations are typically indirect, and hard to associate to the reaction of the model to real natural images.

Analytic Characterization: Certain models can be characterized analytically. One example is the TV regularizer [10], which has been shown to preserve convex shapes as long as the maximal curvature along their boundary is smaller than their perimeter divided by their area [20]. Another example is sparse representations over multiscale frames (*e.g.*, wavelets [23], bandlets [24], curvelets [25], etc.). For instance, contourlets have been shown to provide optimally sparse representations for objects that are piecewise smooth and have smooth boundaries [26] (*i.e.*, functions that are C^2 except for discontinuities along C^2 curves). However, general image priors (especially trained models), are extremely difficult to analyze mathematically.

Patch Based Models: Many parametric models have been used for small image patches, including independent component analysis (ICA) [27], products of experts [28], Gaussian mixture models (GMMs) [1], sparse representation over some dictionary [15], and more. Those models are usually visualized by plotting the basic elements which comprise them. Namely, the independent components in ICA, the dictionary atoms in sparse representations, the top eigenvectors of the Gaussians' covariances in GMM, etc.

Markov Random Fields: These models use Gibbs distributions over filter responses [13, 14, 29–31]. The filters (as well as their potentials) are typically learned from a collection of training images. Those priors can be visualized by drawing samples from the learned model using Markov-chain Monte Carlo (MCMC) simulation [29]. Another common practice is to plot the learnt filters. However, as discussed in [32], those filters are often nonintuitive and difficult to interpret. Indeed, as we show in Sect. 3, our visualization reveals certain geometric preferences of the MRF models [14, 22, 31], which were not previously pointed out.

Deep Networks: These architectures are widely used in image classification, but are also gaining increasing popularity in low-level vision tasks, including in denoising [4], super-resolution [33], and blind deblurring [34]. Visualizing feature activities at different layers has been studied mainly in the context of convolutional networks, and was primarily used to interpret models trained for classification [35, 36]. Features in the first layer typically resemble localized Gabor filters at various orientations, while deeper layers capture structures with increasing complexity.

Patch Recurrence: Patch recurrence is known as a dominant property of natural images. A technique for revealing and modifying variations between repeating structures in an image was recently presented in [37]. This method determines how images should be deformed so as to increase the patch repetitions within them. Although presented in the context of image editing, this method can in fact be viewed as a special case of our proposed approach, where the prior being visualized enforces patch-recurrence within the image. Here, we use the same concept, but to visualize *arbitrary* image priors.

In contrast to previous approaches, which visualize filters, atoms, or other building blocks of the model, our approach rather visualizes the model’s effect on images. As we illustrate, in many cases this visualization is significantly more informative.

2 Algorithm

Suppose we are given a probability model $p(x)$ for natural images. To visualize what geometric properties this model captures, our approach is to determine how images should be deformed so that they become more likely under this model. That is, for any input image y , we seek an idealized version $x \approx \mathcal{T}\{y\}$, for some piecewise-smooth deformation \mathcal{T} , such that $\log p(x)$ is maximal. More specifically, we define the idealizing deformation \mathcal{T} as the solution to the optimization problem

$$\arg \min_{x, \mathcal{T}} \underbrace{-\log p(x)}_{\text{log-prior}} + \underbrace{\lambda \Phi(\mathcal{T})}_{\text{smoothness}} + \underbrace{\frac{1}{2\sigma^2} \|\mathcal{T}\{y\} - x\|^2}_{\text{fidelity}}. \quad (1)$$

The log-prior term forces the image x to be highly plausible under the prior $p(x)$. The smoothness term regularizes the deformation \mathcal{T} to be piecewise smooth. Finally, the fidelity term ensures that the deformed (idealized) input image $\mathcal{T}\{y\}$ is close to x . The parameters σ and λ control the relative weights of the different terms, and as we show in Sect. 2.2, can be used to control the scales of features captured by the visualization.

We use nonparametric deformations, so that the transformation \mathcal{T} is defined as

$$\mathcal{T}\{y\}(\xi, \eta) = y(\xi + u(\xi, \eta), \eta + v(\xi, \eta)) \quad (2)$$

for some flow field (u, v) . We define the smoothness term to be the robust penalty

$$\Phi(\mathcal{T}) = \iint \sqrt{\|\nabla u(\xi, \eta)\|^2 + \|\nabla v(\xi, \eta)\|^2 + \varepsilon^2} d\xi d\eta, \quad (3)$$

where $\nabla = (\frac{\partial}{\partial \xi}, \frac{\partial}{\partial \eta})$ and ε is a small constant. This penalty is commonly used in the optical flow literature [38] and is known to promote smooth flow fields while allowing for sharp discontinuities at objects boundaries.

To solve the optimization problem (1), we use alternating minimization. Namely, we iterate between minimizing the objective w.r.t. the image x while holding the deformation \mathcal{T} fixed, and minimizing the objective w.r.t. \mathcal{T} while holding x fixed.

x -step: The smoothness term in (1) does not depend on x , so that this step reduces to

$$\arg \min_x \frac{1}{2\sigma^2} \|\mathcal{T}\{y\} - x\|^2 - \log p(x). \quad (4)$$

This can be interpreted as computing the maximum a-posteriori (MAP) estimate of x from a “noisy signal” $\mathcal{T}\{y\}$, assuming additive white Gaussian noise with

```

Input: Image  $y$ , denoising function Denoise()
Output: Idealizing transformation  $\mathcal{T}$ , idealized image  $y^{\text{GEM}}$ 
Initialize  $\mathcal{T}^0$  to be the identity mapping.
for  $k = 1, \dots, K$  do
   $x^k \leftarrow \text{Denoise}(\mathcal{T}^{k-1}\{y\})$  /* tuned for noise level  $\sigma$  */
   $\mathcal{T}^k \leftarrow \text{OpticalFlow}(y, x^k)$  /*with regularization parameter  $2\lambda\sigma^2$  */
end
 $\mathcal{T} \leftarrow \mathcal{T}^K$ 
 $y^{\text{GEM}} \leftarrow \mathcal{T}\{y\}$ 

```

Algorithm 1. Geometric prior visualization.

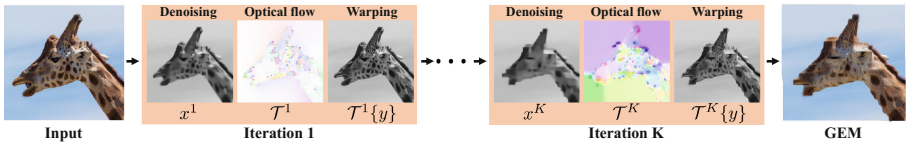


Fig. 3. Schematic illustration of the algorithm. In each iteration, the current corrected image $\mathcal{T}\{y\}$ is “denoised” to obtain an updated image x . Then, the deformation \mathcal{T} is updated to be that which best maps the input y to the new x . This results in a new corrected image $\mathcal{T}\{y\}$. The iterations are shown for the FoE model [14]. Photo courtesy of Mickey Weidenfeld.

variance σ^2 . Thus, x is obtained by “denoising” the current $\mathcal{T}\{y\}$ using the prior $p(x)$.

\mathcal{T} -step: The log-likelihood term in (1) does not depend on \mathcal{T} , so that this step boils down to solving

$$\arg \min_{\mathcal{T}} \|\mathcal{T}\{y\} - x\|^2 + 2\lambda\sigma^2 \cdot \Phi(\mathcal{T}). \quad (5)$$

This corresponds to computing the optical flow between the current image x and the input image y , where the regularization weight is $2\lambda\sigma^2$. To solve this problem we use the iteratively re-weighted least-squares (IRLS) algorithm proposed in [39] (using an L_2 data-term in place of their L_1 term).

Therefore, as summarized in Algorithm 1, our algorithm iterates between denoising the current deformed image, and warping the input image to match the denoised result. Intuitively, when the denoiser is applied on the image, it modifies it to be more plausible according to the prior $p(x)$. This modification introduces slight deformations, among other effects. The role of the optical flow stage is to capture only the geometric modifications, which are those we wish to study. This process is illustrated in Fig. 3.

Note that typical optical flow methods work coarse-to-fine to avoid getting trapped in local minima (the flow computed in each level is interpolated to provide an initialization for the next level). In our case, however, this is not needed because the flow changes very slowly between consecutive iterations of

Algorithm 1. Thus, in each iteration, we simply use the flow from the previous iteration as initialization.

2.1 Alternative Interpretation: Geometric Eigen-Modes

Our discussion so far assumed generative models for whole images. However, many image enhancement algorithms do not explicitly rely on such probabilistic models. Some methods only model the local statistics of small neighborhoods (patches), either by learning from an external database [1], or by relying on the recurrence of patches within the input image itself [16, 17]. Other approaches are discriminative [4], directly learning the desired mapping from input degraded images to output clean images. In all these cases, there is no explicit definition of a probability density function $p(x)$ for whole images, so that the optimization problem (1) is not directly applicable. Nevertheless, note that Algorithm 1 can be used even in the absence of a probability model $p(x)$, as all it requires is the availability of a denoising algorithm. To understand what Algorithm 1 computes when the denoising does not correspond to MAP estimation, it is insightful to examine how the flow \mathcal{T} evolves along the iterations.

Collecting the two steps of Algorithm 1 together, we see that the deformation evolves as $\mathcal{T}^{k+1} = \text{OpticalFlow}(y, \text{Denoise}(\mathcal{T}^k\{y\}))$. Therefore, the algorithm converges once the transformation \mathcal{T} satisfies

$$\mathcal{T} = \text{OpticalFlow}(y, \text{Denoise}(\mathcal{T}\{y\})). \quad (6)$$

This implies that after convergence, denoising $\mathcal{T}\{y\}$ does not introduce geometric deformations anymore. In other words, the output $y^{\text{GEM}} = \mathcal{T}\{y\}$ has the same geometry as its denoised version $\text{Denoise}(y^{\text{GEM}})$. To see this, note that condition (6) states that the image $\text{Denoise}(y^{\text{GEM}})$ is related to y by the deformation \mathcal{T} . But, recall that the image y^{GEM} itself is also related to y by the deformation \mathcal{T} . This is illustrated in Fig. 4.

From the discussion above we conclude the image y^{GEM} produced by our algorithm has the property that *its geometry is not altered by the denoiser*. We therefore call y^{GEM} a *Geometric Eigen-Mode (GEM)* of the prior, associated with image y . Because GEMs are not geometrically modified by the denoiser, the local geometric structures seen in a GEM are precisely those structures which are best preserved by the denoiser. This makes GEMs very informative for studying the geometric preferences of image priors.

2.2 Controlling the Visualization Strength

Recall that the parameters λ and σ control the relative weights of the three terms in Problem¹ (1). To tune the strength of the visualization, we can vary the weight of the log-prior term, which affects the extent to which the ‘idealized’

¹ Strictly speaking, this interpretation is valid only if our denoiser performs MAP estimation. However, the intuition is the same also for arbitrary denoisers.

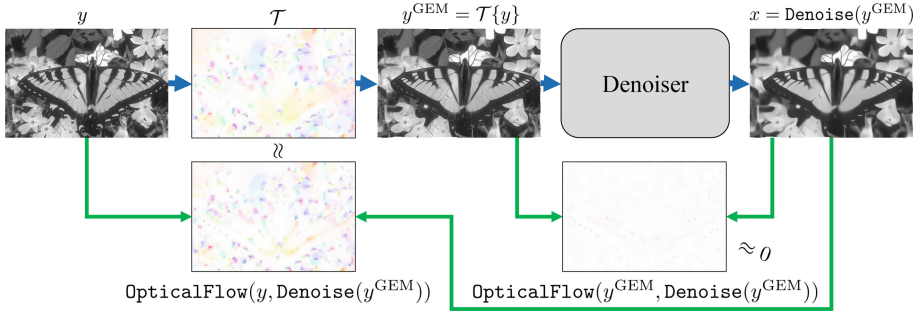


Fig. 4. Denoising a GEM does not change its geometry. The GEM y^{GEM} is obtained by warping the image y with the ‘idealizing’ flow field \mathcal{T} . ‘Denoising’ y^{GEM} , results in an image with the same geometry as y^{GEM} itself. That is, the optical flow between $\text{Denoise}(y^{\text{GEM}})$ and y^{GEM} is zero, and optical flow between $\text{Denoise}(y^{\text{GEM}})$ and y is equal to \mathcal{T} (like the transformation between y^{GEM} itself and y). The results are shown for the multi-layer perceptron (MLP) model [4].

image complies with the prior. This requires varying σ while keeping the product $\lambda\sigma^2$ fixed. Figure 5 shows BM3D-GEMs with several different strengths. As we increase the weight of the log-prior term, smaller and smaller features get deformed so that the prior is better satisfied. This effect is clearly seen in the small arcs, the mandrill’s pupils, and the delicate textures on the mandrill’s fur.

3 Experiments

We used our algorithm on images from [40, 41] and from the Web to study a variety of popular priors [1, 4, 10, 14–17, 22, 31]. Some denoising methods work only on grayscale images. So, for fair comparison, we always determined the idealizing deformation based on the grayscale version of the input image, and then used this deformation to warp the color image itself. In all our experiments we used 50 iterations, $\sigma = 25/50$ and λ in the range $[0.5 \times 10^{-4}, 3 \times 10^{-4}]$ (for gray values in the range $[0, 255]$). Some denoisers do not accept σ as input, like nonlocal means and TV. We tuned those methods’ parameters to perform best in the task of removing noise of variance σ^2 from noisy images.

Figure 6 shows visualization results for BM3D [17], FoE [14], EPLL [1] and TV [10]. As can be seen, common to all these models is that they prefer large structures over small ones. Indeed, note how the small yellow spots on the butterfly, the small arcs in the colosseum, the small black spots on the Dalmatians, and the small white spots on the owl, are all removed in the idealization process (the flow shrinks them until they disappear). The remaining large structures, on the other hand, are distorted quite differently by each of the models.

BM3D [17] is an internal model, which relies on comparisons between patches within the image. As can be seen in Fig. 6, BM3D clearly prefers straight edges connected at sharp corners. Moreover, it favors textures with straight thin

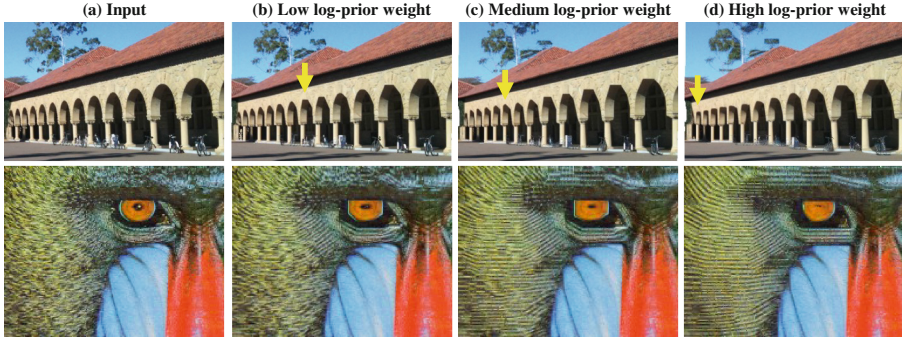


Fig. 5. Controlling the visualization strength. (a) Input images Arcs and Mandril. (b)–(d) BM3D-GEMs with varying strengths, obtained by tuning the log-prior weight in Problem (1). The effect is obtained by increasing σ while decreasing λ so that the product $\lambda\sigma^2$ is kept fixed. We used $\sigma = 20, 30, 50$ in (b), (c), (d), and $\lambda\sigma^2 = 0.128$. As the log-prior weight increases, smaller structures get deformed (*e.g.*, the small arcs and the mandril’s pupils and fur).

threads (see the owl’s head). This can be attributed to the fact that the patch repetitions in those structures are strong. In fact, as we show in Fig. 7, straight edges and sharp corners are also favored by other internal patch-recurrence models, including internally-trained KSVD [15] and the cross-scale patch recurrence prior of [19].

The FoE model [14] expresses the probability of natural images in terms of filter responses. As can be seen in Fig. 6, FoE resonates to straight *axis-aligned* edges connected at right-angle corners. This surprising behavior cannot be predicted by examining the models’ filters, and to the best of our knowledge, was not reported in the past. Note that FoE is an external model that was trained on a collection of images [41]. Therefore, an interesting question is whether its behavior is associated to the statistics of natural images, or rather to some limitation of the model. A partial answer can be obtained by examining the visualizations of EPLL [1], another external model which was trained on the same image collection [41]. As observed in Fig. 6, EPLL also has a preference to straight edges, but its bias towards horizontal and vertical edges is much weaker than that of FoE (a small bias can be noticed on the butterfly’s wings, on the flowers behind the butterfly, and on the Dalmatians’ spots). This suggests that the excessive tendency of FoE to axis-aligned structures is rather related to a limitation of the model, as we further discuss below. We also note that, unlike FoE, the optimal shapes of corners in EPLL are rather round.

Finally, as seen in Fig. 6, the TV prior exhibits a very different behavior. As opposed to all other priors, which prefer straight edges over curved ones, TV clearly preserves curved edges as long as their curvature is not too large. This phenomenon has been studied analytically in [20].

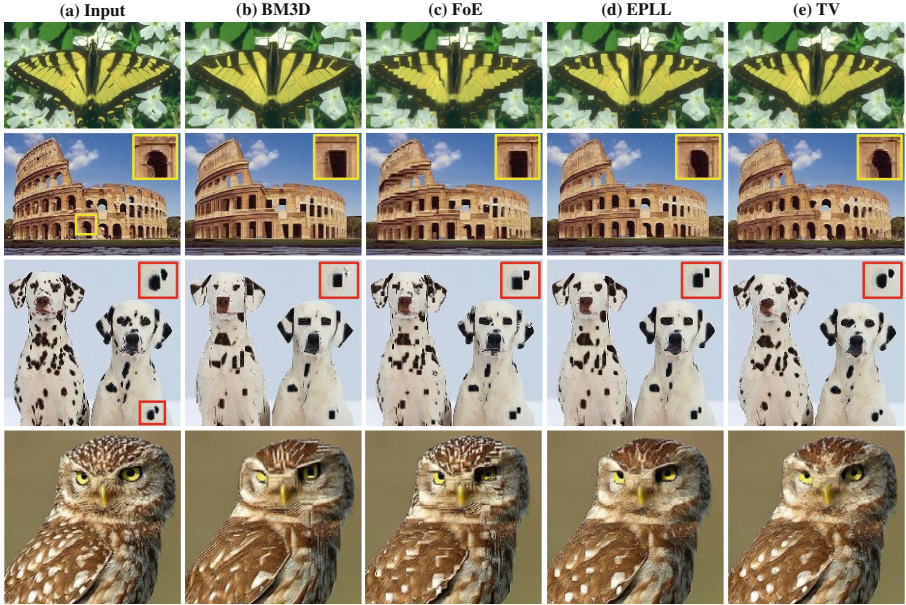


Fig. 6. Visualizing popular image priors. (a) Input images Flower, Colosseum, Dalmatians, and Owl. (b)–(e) Geometric idealization w.r.t. to the BM3D [17], FoE [14], EPLL [1] and TV [10] priors with $\sigma = 50$ and $\lambda = 0.7 \times 10^{-4}$. Note how different elementary structures are preferred by each of the models.

Internal Models: We next compare between several internal models, which rely on the tendency of patches to repeat within and across scales in natural images [42]. Figure 7 shows visualizations for four such methods: BM3D [17], KSVD [15] (trained internally on the input image), the cross-scale patch recurrence model² of [19], and NLM [16]. As can be seen, the GEMs of all these priors have increased redundancy: Edges are deformed to be straighter, stripes are deformed to have constant widths, etc. However, close inspection also reveals interesting differences between the GEMs. Most notably, the NLM method seems to reduce the curvature of edges, but does not entirely straightens them. This may be caused by the fact that it uses a rather localized search window for finding similar patches (15×15 pixels in this experiment). Another noticeable phenomenon, is the thin straight threads appearing in the cross-scale patch recurrence visualization. Those structures are locally self-similar (namely, they look the same at different scales of the image), and are thus preserved by this prior.

External Models: While internal models share a lot in common, external methods exhibit quite diverse phenomena. Figure 8 shows visualizations for several

² This model was presented in [19] in the context of blind deblurring. To use for denoising, we removed the blur-kernel estimation stage and forced the kernel to be a delta function.

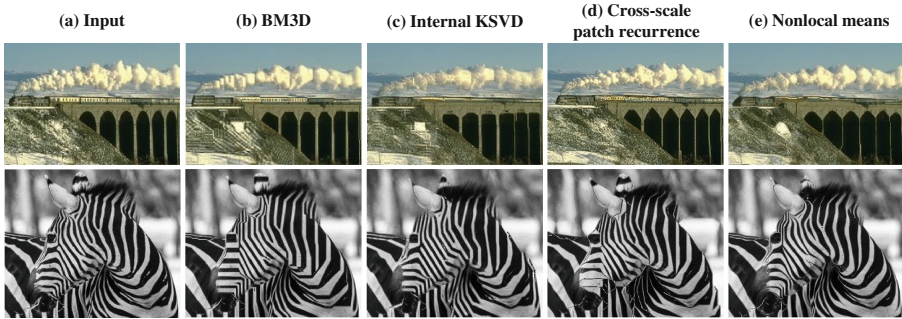


Fig. 7. Comparing internal image models. (a) Input images Train and Zebra (courtesy of Mickey Weidenfeld). (b)–(e) Geometric idealization w.r.t. the BM3D [17], internal KSVD [15], cross-scale patch recurrence [19], and nonlocal means [16] models using $\sigma = 25$ and $\lambda = 2 \times 10^{-4}/3.6 \times 10^{-4}$ for Train/Zebra.

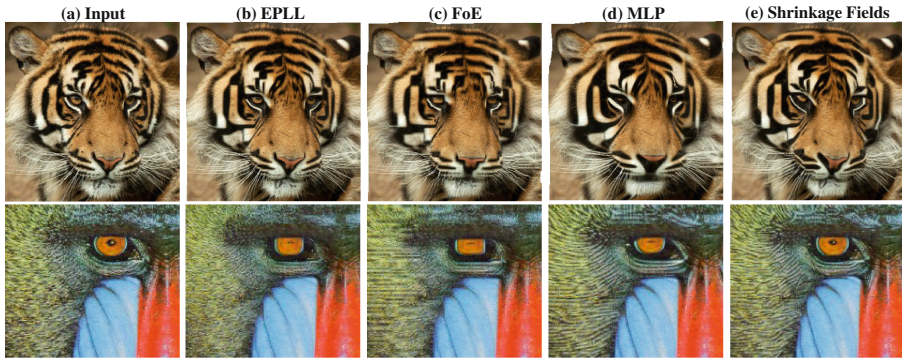


Fig. 8. Comparing external image models. (a) Input images Tiger and Mandril. (b)–(e) Geometric idealization w.r.t. the EPLL [1], FoE [14], multi-layer perceptron (MLP) [4], and Shrinkage Fields [22] models with $\sigma = 25$ and $\lambda = 2 \times 10^{-4}$.

external models, which were all trained on the same dataset [41]: EPLL [1], FoE [14], multi-layer perceptron (MLP) [4], and Shrinkage Fields [22] (an MRF-based model with 7×7 filters). As can be seen, all these models seem to prefer edges with small curvatures. However, apart for FoE, none of them prefers sharp corners. Moreover, the typical shapes of the optimal low-curvature edges differ substantially among these methods. An additional variation among external methods, is that they resonate differently to textures, as can be seen on the mandril’s fur. In the EPLL GEM, the fur is deformed to look smoother, while in all other GEMs, the fur is deformed to exhibit straight strokes.

MRF Models: As mentioned above, the FoE model has a surprising preference to straight axis-aligned edges, significantly more than other external methods trained on the same dataset. This suggests that the FoE model either has limited representation power (*e.g.*, due to the use of 5×5 filters as opposed to the

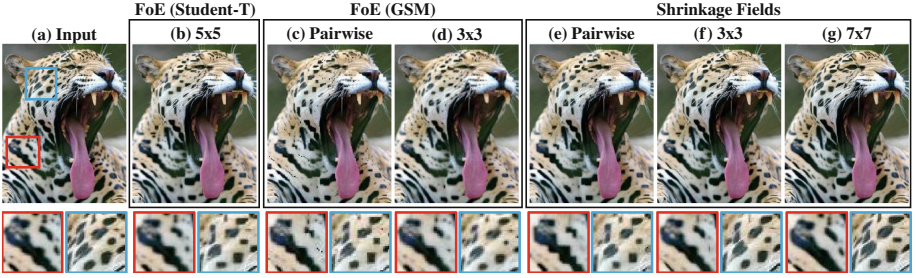


Fig. 9. Comparing MRF image models. (a) Input Jaguar image. (b) GEM of the FoE model with Student-T potentials [14]. (c)–(d) GEMs of FoE model with GSM potentials [31], (e)–(g) GEMs of the Shrinkage Fields model [22]. In all cases $\sigma = 25$ and $\lambda = 2 \times 10^{-4}$.

8×8 patches used in EPLL, or due to the use of Student-T clique potentials), or the learning procedure has converged to a sub-optimal solution. To study this question, Fig. 9 compares the FoE model with [31], an MRF model with Gaussian scale mixture (GSM) clique potentials, and with Shrinkage Fields [22], a discriminative approach which is roughly based on a cascade of several MRF models. The Shrinkage Fields architecture allows efficient training with far larger image crops, than what is practically possible in the FoE model. As can be seen, when using pairwise cliques (horizontal and vertical derivatives), the GSM MRF and Shrinkage Fields also tend to prefer axis-aligned edges. However, this tendency decreases as the filter sizes are increased. With 3×3 filters, in both the GSM MRF and Shrinkage Fields this behavior is already weaker than in the 5×5 FoE model. And for Shrinkage Fields with 7×7 filters, this phenomenon does not exist at all. We confirm this observation in denoising experiments below. While FoE and Shrinkage Fields differ in a variety of aspects (not only the choice of filter sizes), our experiment suggests that MRF models *can* achieve a decent degree of rotation invariance, even with small filters. However, this seems to require large training sets to achieve without intervention. Note that imposing rotation invariance on the filters, has been shown to be beneficial in [32].

3.1 Denoising Experiments

The geometric preferences revealed by our visualizations are very hard, if not impossible, to visually perceive by the naked eye in conventional image recovery experiments on natural images (*e.g.*, denoising, deblurring, super-resolution, etc.). This raises the question: To what extent do these geometric preferences affect the recovery error in such tasks? To study this question, we performed several denoising experiments.

Denoising GEMs: We begin by examining how much easier it is for denoising methods to remove noise from the GEM of an image, than from the image itself. Intuitively, since GEMs contain structures that best conform to the prior,

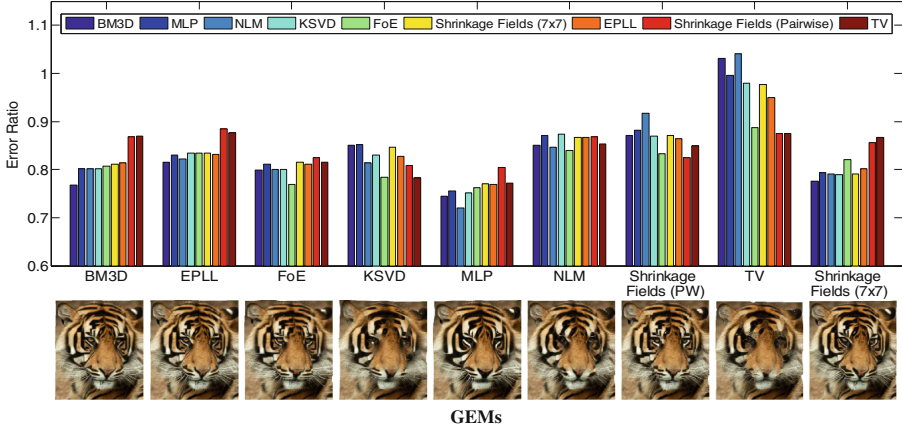


Fig. 10. Denoising GEMs. We added noise to the GEMs corresponding to various priors, and then denoised each of them using various denoising methods. For each denoiser, we report the ratio between the MSE it achieves in denoising the GEM, and the MSE it achieves in denoising the original image. Each color corresponds to a different denoiser, and each group of bars corresponds to a different GEM.

denoising a GEM should be an easier task. Denote by y_p^{GEM} the GEM of image y according to prior p (e.g., $p \in \{\text{‘BM3D’}, \text{‘MLP’}, \dots\}$). We define the *error ratio*

$$r_{p,q}(y) = \frac{\text{MSE}_q(y_p^{\text{GEM}})}{\text{MSE}_q(y)}, \quad (7)$$

where $\text{MSE}_q(y_p^{\text{GEM}})$ and $\text{MSE}_q(y)$ denote the mean square errors (MSEs) attained in recovering the images y_p^{GEM} and y , respectively, from their noisy versions, based on prior q . An error ratio smaller than 1 indicates that recovering y_p^{GEM} with prior q leads to better MSE than recovering y itself with prior q .

Figure 10 shows the error ratios attained by 9 different denoising methods (colored bars), on the 9 GEMs of the corresponding priors (groups of bars) for the tiger image of Fig. 8(a). As can be seen, all the denoisers attain an error ratio smaller than 1 on the GEMs corresponding to their prior (namely $r_{p,p}(y) < 1$ for all p). Moreover, almost all the denoisers attain error ratios smaller than 1 also on the GEMs corresponding to *other* priors³. This suggests that the geometric structures that are optimal for one prior are usually quite good also for other priors.

This experiment further highlights several interesting behaviors. BM3D and NLM perform very poorly on the TV-GEM. This illustrates that an image with low total-variation (the TV-GEM) does not necessarily have strong patch rep-

³ Note that some denoisers perform better on the GEMs of other priors than on their own GEM. This is because GEMs are not optimized to minimize the MSE in denoising tasks. Their construction also takes into account a penalty on the deformation smoothness.

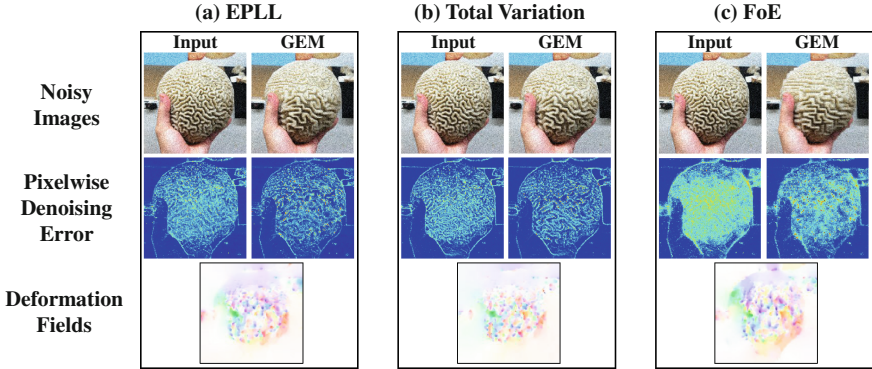


Fig. 11. Pixelwise RMSE. We compare between the pixelwise RMSE (averaged over 50 noise realizations) attained in denoising an image and its GEM. Results are shown alongside the deformation field for (a) EPLL [1], (b) Total variation [10] and (c) FoE [14]. As can be seen, a significant RMSE improvement is achieved in regions which undergo a large deformation.

ethods (as required by the BM3D and NLM denoisers). Shrinkage Fields with pairwise cliques and TV perform very similarly on all the GEMs, and quite differently from all other methods. This may be associated to the fact that they are the only priors based on derivatives. Another distinctive group is MLP, Shrinkage Fields (7×7) and EPLL, which perform similarly on all the GEMs. Common to these methods, is that they are all based on external models trained on the same dataset.

Pixelwise MSE: We next visualize which pixels in a GEM contribute the most to the improved ability to denoise it. Figure 11 shows the pixelwise root-MSE (RMSE) attained in denoising the Brain Coral image and its GEM (using the GEM’s prior), averaged over 50 noise realizations. As can be seen, the largest RMSE improvement occurs at regions which are strongly deformed. Those regions are precisely the places which did not comply with the model initially, and were ‘corrected’ in the GEM.

Rotation Invariance: Our visualizations in Figs. 6, 8, and 9, revealed an interesting preference to axis aligned edges for some of the priors (especially FoE). To verify whether our observations are correct, we plot in Fig. 12 the RMSE that different methods attain in denoising images of rotated squares. As predicted by our visualizations, among external models, the FoE prior indeed has the least degree of rotation invariance, followed by Shrinkage Fields with pairwise cliques. The RMSE of these two methods drops significantly as the angle of the square approaches 0. It can be seen that EPLL also has a slight tendency to axis-aligned edges, while Shrinkage Fields (7×7) is almost entirely indifferent to the square’s angle. These behaviors align with our conclusions from Figs. 8 and 9. We note, however, that MLP also seems to perform slightly better in denoising axis-aligned squares, a behavior that we could not clearly see in the

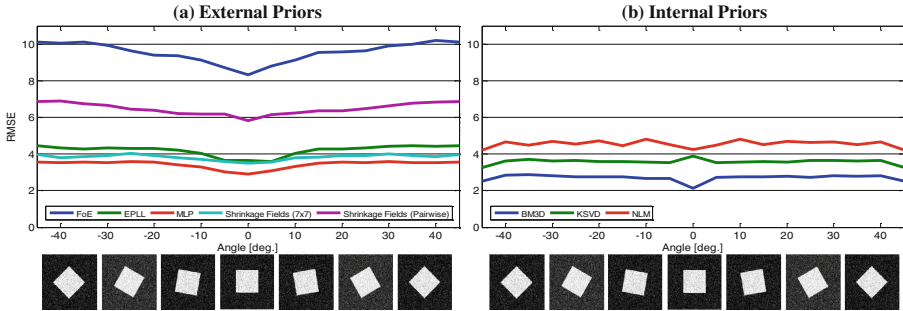


Fig. 12. Rotation invariance. The RMSE attained by various denoising methods in the task of removing noise from a noisy square, as a function of the square’s angle. (a) Methods based on external priors. (b) Methods based on internal priors.

GEM of Fig. 8. The internal models, shown in Fig. 12(b), are almost completely insensitive to the square’s angle, which aligns with the behaviors we observed in the GEMs of Fig. 7. The singular behaviors at angles 0 and 45 are related to the fact that these are the only two angles in which the rotated square does not involve interpolation artifacts.

4 Conclusions

We presented an algorithm for visualizing the geometric preferences of image priors. Our method determines how an image should be deformed so as to best comply with a given image model. Our approach is generic and can be used to visualize arbitrary priors, providing a useful means to study and compare between them. Applying our method on several popular image models, we found various interesting behaviors that are impossible to see using any other visualization technique. Although we demonstrated our approach in the context of visualizing *geometric* properties of image models, our framework can be easily generalized to other types of transformations (*e.g.*, color mappings). This only requires replacing the optical-flow stage in our algorithm accordingly. Our visualizations can be used to analyze failures and successes of image models in specific settings, and may thus help to identify potential model improvements, which are of great importance in image enhancement tasks.

Acknowledgements. This research was supported in part by the Ollendorff Foundation and the Horev and Alon Fellowships.

References

1. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: IEEE International Conference on Computer Vision, pp. 479–486 (2011)

2. Levin, A.: Blind motion deblurring using image statistics. In: *Advances in Neural Information Processing Systems*, pp. 841–848 (2006)
3. Elad, M., Starck, J.L., Querre, P., Donoho, D.L.: Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Appl. Comput. Harmonic Anal.* **19**(3), 340–358 (2005)
4. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: can plain neural networks compete with BM3D? In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2392–2399 (2012)
5. Zontak, M., Mosseri, I., Irani, M.: Separating signal from noise using patch recurrence across scales. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1195–1202 (2013)
6. Sun, J., Sun, J., Xu, Z., Shum, H.Y.: Image super-resolution using gradient profile prior. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8 (2008)
7. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **19**(11), 2861–2873 (2010)
8. Levin, A., Zomet, A., Weiss, Y.: Learning how to inpaint from global image statistics. In: *Proceedings of Ninth IEEE International Conference on Computer Vision*, pp. 305–312 (2003)
9. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 417–424 (2000)
10. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Phys. D: Nonlinear Phenom.* **60**(1), 259–268 (1992)
11. Donoho, D.L.: De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **41**(3), 613–627 (1995)
12. Portilla, J., Strela, V., Wainwright, M.J., Simoncelli, E.P.: Image denoising using scale mixtures of Gaussians in the wavelet domain. *IEEE Trans. Image Process.* **12**(11), 1338–1351 (2003)
13. Zhu, S.C., Mumford, D.: Prior learning and Gibbs reaction-diffusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(11), 1236–1250 (1997)
14. Roth, S., Black, M.J.: Fields of experts: a framework for learning image priors. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 860–867 (2005)
15. Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.* **15**(12), 3736–3745 (2006)
16. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 60–65 (2005)
17. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising with block-matching and 3D filtering. In: *Electronic Imaging*, p. 606414 (2006)
18. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: *International Conference on Computer Vision (ICCV)* (2009)
19. Michaeli, T., Irani, M.: Blind deblurring using internal patch recurrence. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) *ECCV 2014*. LNCS, vol. 8691, pp. 783–798. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-10578-9_51](https://doi.org/10.1007/978-3-319-10578-9_51)
20. Bellettini, G., Caselles, V., Novaga, M.: The total variation flow in \mathbb{R}^N . *J. Differ. Eqn.* **184**(2), 475–525 (2002)
21. Freedman, G., Fattal, R.: Image and video upscaling from local self-examples. *ACM Trans. Graph. (TOG)* **30**(2), 12 (2011)

22. Schmidt, U., Roth, S.: Shrinkage fields for effective image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2774–2781 (2014)
23. Mallat, S.: A wavelet tour of signal processing: the sparse way (2008)
24. Candes, E.J., Donoho, D.L.: Curvelets: a surprisingly effective nonadaptive representation for objects with edges. Technical report, DTIC Document (2000)
25. Starck, J.L., Candès, E.J., Donoho, D.L.: The curvelet transform for image denoising. *IEEE Trans. Image Process.* **11**(6), 670–684 (2002)
26. Candès, E.J., Donoho, D.L.: New tight frames of curvelets and optimal representations of objects with piecewise C^∞ singularities. *Commun. Pure Appl. Math.* **57**(2), 219–266 (2004)
27. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. *Neural Netw.* **13**(4), 411–430 (2000)
28. Hinton, G.E.: Products of experts. In: Ninth International Conference on Artificial Neural Networks, vol. 1, pp. 1–6 (1999)
29. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741 (1984)
30. Freeman, W.T., Pasztor, E.C., Carmichael, O.T.: Learning low-level vision. *Int. J. Comput. Vision* **40**(1), 25–47 (2000)
31. Schmidt, U., Gao, Q., Roth, S.: A Generative perspective on MRFs in low-level vision. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1751–1758. IEEE (2010)
32. Weiss, Y., Freeman, W.T.: What makes a good model of natural images? In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2007)
33. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-10593-2_13](https://doi.org/10.1007/978-3-319-10593-2_13)
34. Schuler, C.J., Hirsch, M., Harmeling, S., Schölkopf, B.: Learning to Deblur. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(7), 1439–1451 (2015)
35. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Heidelberg (2014). doi:[10.1007/978-3-319-10590-1_53](https://doi.org/10.1007/978-3-319-10590-1_53)
36. Zeiler, M.D., Krishnan, D., Taylor, G.W., Fergus, R.: Deconvolutional networks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2528–2535 (2010)
37. Dekel, T., Michaeli, T., Irani, M., Freeman, W.T.: Revealing and modifying non-local variations in a single image. *ACM Trans. Graph. (TOG)* **34**(6), 227 (2015)
38. Brox, T., Bruhn, A., Papenberger, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004). doi:[10.1007/978-3-540-24673-2_3](https://doi.org/10.1007/978-3-540-24673-2_3)
39. Liu, C.: Beyond pixels: exploring new representations and applications for motion analysis. Ph.D. thesis, Citeseer (2009)
40. Rubinstein, M., Gutierrez, D., Sorkine, O., Shamir, A.: A Comparative study of image retargeting. *ACM Trans. Graph.* **29**(6) 160:1–160:10 (2010). (Proc. SIGGRAPH Asia)

41. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A Database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings of the 8th International Conference on Computer Vision, vol. 2. 416–423, July 2001
42. Zontak, M., Irani, M.: Internal statistics of a single natural image. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 977–984 (2011)