Supplemental Material for the Paper
"Projection onto the Manifold of Elongated Structures for Accurate Extraction"

Amos Sironi[1] Vincent Lepetit[1,2] Pascal Fua[1]
[1]CVLab, EPFL, Lausanne, Switzerland, {firstname.lastname}@epfl.ch
[2]TU Graz, Graz, Austria, lepetit@icg.tugraz.at

In Section A of this appendix we formalize the concepts introduced in Section 4.2 of our submission and prove the equivalence stated there. Then, in Section B we describe some implementation details of our method. Finally, in Section C we present additional results that did not fit in the main submission because of space limitations.

A. Equivalence of $\Pi_{D\to N}(X)$ and $\Pi_N(X)$

In this section we prove that the output $\Pi_{D\to N}(X)$ of our method is equivalent to the projection $\Pi_N(X)$ of the score map X into a manifold of admissible ground truth $\mathcal{M}_N$.

A.1. Notation

We start by defining the notation used in this section and recalling the definitions given in the main submission.

let $I \in \mathbb{R}^N$ be an image containing elongated structures we are interested in extracting.

Let $\{p_i\}_{i=1}^N$ be the grid of pixels on which the images are defined, where for simplicity we assume squared images, $N = \sqrt{N} \times \sqrt{N}$. For a pixel p_i, we denote by $\mathcal{N}_D(p_i)$ the squared neighborhood of pixels centered at p_i of size $D = \sqrt{D} \times \sqrt{D}$.

Let $Y \in \{0,1\}^N$ be the binary ground truth corresponding to image I and let $dY \in \mathbb{R}^N$ be the image obtained by applying function d to every pixel of Y, where d is defined by

$$d(p) = \begin{cases} e^{a(1-\frac{\mathcal{D}_Y(p)}{d_M})} - 1 & \text{if } \mathcal{D}_Y(p) < d_M \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

with $\mathcal{D}_Y$ the Euclidean distance transform of Y, $a > 0$ a constant that controls the exponential decrease rate of d close to the centerline and $d_M > 0$ a threshold value determining how far from a centerline d is set to zero.

We denote by $\mathcal{M}_N$ the manifold of images dY, obtained from all possible ground truth images Y.

We define X the score image obtained by applying the regressor φ of Section 3.1 to every pixel of image I. The projection of X into $\mathcal{M}_N$ is denoted by $\Pi_N(X)$ and it is given by

$$\Pi_N(X) = \underset{dY \in \mathcal{M}_N}{\arg\min} \|dY - X\|^2. \tag{2}$$

As described in Section 4, we approximate $\Pi_N(X)$ by averaging the projections of small patches of X. In order to formalize this, we define for every pixel p_i of image X the corresponding patch x_i of size D centered at p_i, $x_i = X(\mathcal{N}_D(p_i))$. We also denote by $\mathcal{M}_D$ the manifold of all patches of size D extracted from images in $\mathcal{M}_N$.

The projection of x_i into $\mathcal{M}_D$ is given by

$$\Pi_D(x_i) = \underset{y \in \mathcal{M}_D}{\arg\min} \|y - x_i\|^2. \tag{3}$$

The approximated projection $\Pi_{D\to N}(X)$ is defined as

$$\Pi_{D\to N}(X)(p) = \frac{1}{R} \sum_{i: p-p_i \in \mathcal{N}_R(p)} \Pi_D(x_i)(p - p_i), \tag{4}$$

where $R \leq D$ is the size of the neighborhood used for averaging and where we take $\Pi_D(x_i)$ to be centered at zero, with $\Pi_D(x_i)(p-p_i)$ the value of $\Pi_D(x_i)$ at $p-p_i$. In the following, to simplify notations, we will consider the case $R = D$. The generalization to a generic value of R is straightforward.

Let $\{q_d\}_{d=1}^D$ be the $\sqrt{D} \times \sqrt{D}$ grid of pixels on which the patches of size D are defined. For a pixel $p \in \{p_i\}_{i=1}^N$ in the image grid of pixels, we denote with $p^{(l)}$, for $l = 1, \dots, D$ the elements of $\mathcal{N}_D(p)$.

For an image patch x of size D, we say that the *support* of x is $\mathcal{N}_D(p)$ if, between the sets $\{q_d\}_{d=1}^D$ and $\mathcal{N}_D(p)$, there is the one to one correspondence f, given by $q_l = f(p^{(l)})$. In other words, this means that patch x can be thought as centered at image location p. In such case, to simplify our notation, we will write $x(p^{(d)})$ instead of $x(f(p^{(d)}))$.

For example, given the score image X, the image patch $x_i = X(\mathcal{N}_D(p_i))$ centered at p_i, has support $\mathcal{N}_D(p_i)$. In the same way, we will define the support of the projections $\Pi_D(x_i)$ to be $\mathcal{N}_D(p_i)$. In this way, we can rewrite the definition of $\Pi_{D\to N}(X)$ Eq. (5) as

$$\Pi_{D\to N}(X)(p) = \frac{1}{D} \sum_{i:p\in\mathcal{N}_D(p_i)} \Pi_D(x_i)(p). \tag{5}$$

We say that two patches x_i and x_j *overlap* if the intersection of their supports $\mathcal{N}_D(p_i) \cap \mathcal{N}_D(p_j)$ is not empty.

We indicate with $X|_{\mathcal{N}}$ the restriction of an image X to a subset of pixels $\mathcal{N} \subseteq \{p_i\}_i$. For example, using this notation, we can write $x_i = X|_{\mathcal{N}_D(p_i)}$.

Let $\{y_j\}_{j=1}^J$ be a set of patches of size D such that for every j, y_j has support $\mathcal{N}_D(p_j)$. We say that $\{y_j\}_{j=1}^J$ *covers* the pixel grid $\{p_i\}_{i=1}^N$ if

$$\bigcup_j \mathcal{N}_D(p_j) = \{p_i\}_{i=1}^N. \tag{6}$$

This means that for every pixel p_i there exist at least one j such that p_i is in the support of y_j.

Then, for a set of patches $\{y_j\}_{j=1}^J$, such that $\{y_j\}_{j=1}^J$ covers $\{p_i\}_{i=1}^N$, we can define a new image of size N, called the *average image of patches* $\{y_j\}_{j=1}^J$ and denoted by $\bigsqcup_{j=1}^J y_i$, by averaging the the patches y_j on their supports. More precisely, for every pixel $p \in \{p_i\}_i$, the value of the average image at pixel p is given by

$$\left(\bigsqcup_{j=1}^J y_j\right)(p) = \frac{1}{C_p} \sum_{j:\ p\in\mathcal{N}_D(p_j)} y_j(p) \tag{7}$$

Where the normalization constant C_p is equal to the number of elements in the sum and in general it depends on the pixel location p. Notice that (7) is well defined for every p because of (6).

In other words, for a pixel p, the value of $\bigsqcup_{j=1}^J y_j$ in p is given by the average of the values $y_j(p)$, for those y_j with support containing p. For example, in the case of image patches $x_i = X(\mathcal{N}_D(p_i))$, since $x_i(p) = x_j(p)$ for all $p \in \mathcal{N}_D(p_i) \cap \mathcal{N}_D(p_j)$, we have

$$X = \bigsqcup_{i=1}^N x_i. \tag{8}$$

Notice that in this case we have $C_p = D$ for every p.

With this notations, using the fact that the support of $\Pi_D(x_i)$ is $\mathcal{N}_D(p_i)$, we can write the approximated projection $\Pi_{D\to N}(X)$ as

$$\Pi_{D\to N}(X) = \bigsqcup_{i=1}^N \Pi_D(x_i). \tag{9}$$

Also in this case the normalization constant in Eq. (7) is $C_p = D$ for every p.

A.2. Exact Projection

In this section we prove the equivalence $\Pi_{D\to N}(X) = \Pi_N(X)$. We first give two definitions that will be used as hypothesis in the following Theorems.

Definition 1. *We say that $\mathcal{M}_D$ is* complete *in $\mathcal{M}_N$ if the following holds:*

$$Y \in \mathcal{M}_N \iff \exists \{y_i\}_{i=1}^N \subseteq \mathcal{M}_D \text{ such that } Y = \bigsqcup_{i=1}^N y_i, \tag{10}$$

where, $\forall i$, y_i has support $\mathcal{N}_D(p_i)$ and $\forall i, j$, $y_i(p) = y_j(p)$ for all $p \in \mathcal{N}_D(p_i) \cap \mathcal{N}_D(p_j)$.

This is property (i) given in Section 4.2 of our main submission. It means that the training set of patches $\mathcal{M}_D$ is composed by all admissible ground truth patches and that averaging patches that coincide in the intersection of their supports, gives an image of $\mathcal{M}_N$.

The following definition formalizes instead hypothesis (ii) of Section 4.2.

Definition 2. *Let us consider the set of all projections* $\{\Pi_D(x_i)\}_{i=1}^N$ *of all patches of size* D *of an image* X*. We say that* $\{\Pi_D(x_i)\}_i$ *is* consistent *if, for all* x_i *and* x_j *such that* $\mathcal{N}_D(p_i) \cap \mathcal{N}_D(p_j) \neq \emptyset$*, we have* $\Pi_D(x_i)(p) = \Pi_D(x_j)(p)$ *for all* $p \in \mathcal{N}_D(p_i) \cap \mathcal{N}_D(p_j)$*.*

This means that the projection of two overlapping patches is the same for every pixel in the intersection of their support.

We can now state the following

Theorem 1. *If* $\mathcal{M}_D$ *is complete in* $\mathcal{M}_N$ *and if* $\{\Pi_D(x_i)\}_{i=1}^N$ *is consistent, then*

$$\|X - \Pi_{D\to N}(X)\|^2 = \|X - \Pi_N(X)\|^2. \tag{11}$$

Proof. Since $\Pi_N(X) \in \mathcal{M}_N$ and $\mathcal{M}_D$ is complete, for every patch x_i the restriction of $\Pi_N(X)$ to x_i belongs to $\mathcal{M}_D$, $\Pi_N(X)|_{\mathcal{N}_D(p_i)} \in \mathcal{M}_D$. Then, by (3), we have for all x_i

$$\|\Pi_D(x_i) - x_i\|^2 \leq \|\Pi_N(X)|_{\mathcal{N}(p_i)} - x_i\|^2. \tag{12}$$

Let $\{x_{i_j}\}_{j=1}^K$ be a subset of $\{x_i\}_{i=1}^N$ such that $\mathcal{N}(p_{i_{j_1}}) \cap \mathcal{N}(p_{i_{j_2}}) = \emptyset$ for all $j_1 \neq j_2$ and $X = \bigsqcup_j x_{i_j}$. The subset of patches $\{x_{i_j}\}_{j=1}^K$ is given by a grid of non-overlapping image patches covering the whole image [1].

Since $\{\Pi_D(x_i)\}_i$ is consistent, $\Pi_{D\to N}(X) = \bigsqcup_j \Pi_D(x_{i_j})$. In fact, for the hypothesis of consistency, the patches $\Pi_D(x_i)$ coincide in their intersection. Then,

$$\begin{aligned}\|\Pi_{D\to N}(X) - X\|^2 &= \|\bigsqcup_j \Pi_D(x_{i_j}) - \bigsqcup_j x_{i_j}\|^2 = \\ &= \sum_j \|\Pi_D(x_{i_j}) - x_{i_j}\|^2.\end{aligned} \tag{13}$$

Where the second equality holds since patches x_{i_j} do not overlap. Eq.(13) tells that the distance between images X and $\Pi_{D\to N}(X)$ can be computed by summing the distances between non-overlapping patches $\Pi_D(x_{i_j})$ and x_{i_j}.

Then, from (12) and (13)

$$\begin{aligned}\|\Pi_{D\to N}(X) - X\|^2 &\leq \sum_j \|\Pi_N(X)|_{\mathcal{N}(p_{i_j})} - x_{i_j}\|^2 = \\ &= \|\Pi_N(X) - X\|^2,\end{aligned} \tag{14}$$

where again the second equality follows by the fact that patches x_{i_j} do not overlap.

However, since $\mathcal{M}_D$ is complete and $\{\Pi_D(x_i)\}_i$ is consistent, $\Pi_{D\to N}(X) \in \mathcal{M}_N$. In fact $\Pi_{D\to N}(X)$ is given by the average of the projections $\Pi_D(x_i)$ satisfying definition (10). Therefore, since $\Pi_N(X)$ is defined in (2) as the point of minimum distance to X in $\mathcal{M}_N$, we have

$$\|\Pi_N(X) - X\|^2 \leq \|\Pi_{D\to N}(X) - X\|^2. \tag{15}$$

From (14) and (15) we have the thesis. □

Notice that the thesis of Theorem 1 tells us that the distance between our approximation $\Pi_{D\to N}(X)$ and X is as good as the distance between X and $\Pi_N(X)$. The equivalence of the two projections is given assuming that there is a unique minimum of the distance to X in $\mathcal{M}_N$. This is stated in the following

Corollary 1. *In the hypothesis of Theorem 1, if the function* $F(Y) = \|Y - X\|$ *has a unique minimum in* $\mathcal{M}_N$*, we have*

$$\Pi_N(X) = \Pi_{D\to N}(X). \tag{16}$$

Proof. The thesis follows by (11) and the hypothesis. □

[1] Such a decomposition always exists assuming that we can pad images with zeros.

In real applications only limited training patches are available and the hypothesis of completeness might not be satisfied. Also consistency will not hold in general and projections in the intersection of overlapping patches will not be exactly the same.

However, in next Section we show that by relaxing the hypothesis of Theorem 1 and assuming only approximated projections, we can prove that the error committed by our method is within a certain bound to the optimal solution. This bound is directly related to the error committed by the projections on the patches $\Pi_D(x_i)$ and the size of our training set.

A.3. Approximated Projection

Let $\bar{\mathcal{M}}_D$ the set of all patches y_i of size D of images dY in $\mathcal{M}_N$, assume that $\bar{\mathcal{M}}_D$ is complete in $\mathcal{M}_N$[2] and that the set of available training patches $\mathcal{M}_D$ is included in $\bar{\mathcal{M}}_D$, $\mathcal{M}_D \subseteq \bar{\mathcal{M}}_D$. For a patch x, we denote $\Pi_{\bar{D}}(x)$ the projection of x into $\bar{\mathcal{M}}_D$.

For $\epsilon \geq 0$, we say that $\mathcal{M}_D$ is $\epsilon-$*complete* in $\mathcal{M}_N$, if for all $y \in \mathcal{M}_D$ there exists $\bar{y} \in \bar{\mathcal{M}}_D$ such that $\|y - \bar{y}\|^2 \leq \epsilon$. This means that every ground truth patch $\bar{y}$ is close, up to an error of ε, to an available training patch y. This hypothesis will replace the hypothesis of completeness of Theorem 1.

The hypothesis of consistency will be replaced by the following:

$$\exists \epsilon_1 \geq 0 \text{ s.t. } \forall x_i \ \|\Pi_{D\to N}(X)|_{\mathcal{N}(p_i)} - \Pi_D(x_i)\|^2 \leq \epsilon_1, \tag{17}$$

$$\exists \epsilon_2 \geq 0 \text{ s.t. } \forall x_i \ \|\Pi_N(X)|_{\mathcal{N}(p_i)} - \Pi_{\bar{D}}(x_i)\|^2 \leq \epsilon_2. \tag{18}$$

This means that the restriction to $\mathcal{N}(p_i)$ of the images $\Pi_{D\to N}(X)$ and $\Pi_N(X)$ are close to the projections of patch x_i onto $\mathcal{M}_D$ and $\bar{\mathcal{M}}_D$ respectively. Note that if $\{\Pi_D(x_i)\}_i$ is consistent, then $\epsilon_1 = 0$ and if $\{\Pi_{\bar{D}}(x_i)\}_i$ is consistent, then $\epsilon_2 = 0$.

We now have the following

Theorem 2. *If $\mathcal{M}_D$ is $\epsilon-$complete in $\mathcal{M}_N$ and if* (17) *and* (18) *hold. Then,*

$$-\sqrt{\frac{N}{D}}(\epsilon_1 + \epsilon) \leq \|\Pi_N(X) - X\|^2 - \|\Pi_{D\to N}(X) - X\|^2 \leq \sqrt{\frac{N}{D}}(\epsilon_1 + \epsilon_2). \tag{19}$$

Proof. Let $\{x_{i_j}\}_j$ be a disjoint partition of the image, as in the proof of Theorem 1. Then,

$$\begin{aligned}
\|\Pi_N(X) - X\|^2 &= \sum_j \|\Pi_N(X)|_{\mathcal{N}(p_{i_j})} - x_{i_j}\|^2 \leq \\
&\leq \sum_j \left(\|\Pi_N(X)|_{\mathcal{N}(p_{i_j})} - \Pi_{\bar{D}}(x_{i_j})\|^2 + \|\Pi_{\bar{D}}(x_{i_j}) - x_{i_j}\|^2\right) \leq \\
&\leq \sum_j \left(\epsilon_2 + \|\Pi_{\bar{D}}(x_{i_j}) - x_{i_j}\|^2\right) = \\
&= \sqrt{\frac{N}{D}}\epsilon_2 + \sum_j \|\Pi_{\bar{D}}(x_{i_j}) - x_{i_j}\|^2.
\end{aligned} \tag{20}$$

Where the first inequality is the triangle inequality and the second inequality is given by hypothesis (18).

Since $\mathcal{M}_D \subseteq \bar{\mathcal{M}}_D$, for every x_i $\|\Pi_{\bar{D}}(x_i) - x_i\|^2 \leq \|\Pi_D(x_i) - x_i\|^2$. In fact, $\Pi_{\bar{D}}(x_i)$ is the point of minimum distance to x_i, computed on a larger set $\bar{\mathcal{M}}_D$ compared to $\Pi_D(x_i)$.

Then, Eq.(20) becomes

$$\|\Pi_N(X) - X\|^2 \leq \sqrt{\frac{N}{D}}\epsilon_2 + \sum_j \|\Pi_D(x_{i_j}) - x_{i_j}\|^2. \tag{21}$$

[2] If $\bar{\mathcal{M}}_D$ is not complete, we can extend $\mathcal{M}_N$ by adding to it all images that can be obtained with Eq. (7) from patches in $\bar{\mathcal{M}}_D$ that coincide in their supports.

By using the triangle inequality and hypothesis (17), Eq. (21) becomes

$$
\begin{aligned}
\|\Pi_N(X) - X\|^2 \leq \sqrt{\frac{N}{D}}\epsilon_2 + \sum_j \Big(\|\Pi_D(x_{i_j}) - \Pi_{D\to N}(X)|_{\mathcal{N}_D(p_{i_j})}\|^2 \\
+ \|\Pi_{D\to N}(X)|_{\mathcal{N}_D(p_{i_j})} - x_{i_j}\|^2 \Big) \leq \\
\leq \sqrt{\frac{N}{D}}\epsilon_2 + \sum_j \Big(\epsilon_1 + \|\Pi_{D\to N}(X)|_{\mathcal{N}_D(p_{i_j})} - x_{i_j}\|^2 \Big) = \\
= \sqrt{\frac{N}{D}}(\epsilon_2 + \epsilon_1) + \sum_j \|\Pi_{D\to N}(X)|_{\mathcal{N}_D(p_{i_j})} - x_{i_j}\|^2 = \\
= \sqrt{\frac{N}{D}}(\epsilon_2 + \epsilon_1) + \|\Pi_{D\to N}(X) - X\|^2.
\end{aligned}
\tag{22}
$$

The inequality in (22) proves the right hand side in (19). For the left hand side we proceed analogously.

$$
\begin{aligned}
\|\Pi_{D\to N}(X) - X\|^2 = \sum_j \|\Pi_{D\to N}(X)|_{\mathcal{N}_D(p_{i_j})} - x_{i_j}\|^2 \leq \\
\leq \sum_j \Big(\|\Pi_{D\to N}(X)|_{\mathcal{N}_D(p_{i_j})} - \Pi_D(x_{i_j})\|^2 + \|\Pi_D(x_{i_j}) - x_{i_j}\|^2 \Big) \leq \\
\leq \sqrt{\frac{N}{D}}\epsilon_1 + \sum_j \|\Pi_D(x_{i_j}) - x_{i_j}\|^2.
\end{aligned}
\tag{23}
$$

Where we first used triangle inequality and then hypothesis (17).

Since $\Pi_N(X)|_{\mathcal{N}_D(p_i)} \in \bar{\mathcal{M}}_D$ for all p_i, and since $\mathcal{M}_D$ is $\epsilon-$complete, $\forall p_i$ there exists $y_i \in \mathcal{M}_D$ such that $\|y_i - \Pi_N(X)|_{\mathcal{N}(p_i)}\|^2 \leq \epsilon$. Moreover, for all $y_i \in \mathcal{M}_D$, $\|\Pi_D(x_i) - x_i\|^2 \leq \|y_i - x_i\|^2$. In fact, $\Pi_D(x_i)$ is the point of minimum distance to x_i in $\mathcal{M}_D$. Hence, substituting in (23) we have

$$
\begin{aligned}
\|\Pi_{D\to N}(X) - X\|^2 \leq \sqrt{\frac{N}{D}}\epsilon_1 + \sum_j \|\Pi_D(x_{i_j}) - x_{i_j}\|^2 \leq \\
\leq \sqrt{\frac{N}{D}}\epsilon_1 + \sum_j \|y_{i_j} - x_{i_j}\|^2 \leq \\
\leq \sqrt{\frac{N}{D}}\epsilon_1 + \sum_j \Big(\|y_{i_j} - \Pi_N(X)|_{\mathcal{N}_D(p_{i_j})}\|^2 + \\
+ \|\Pi_N(X)|_{\mathcal{N}_D(p_{i_j})} - x_{i_j}\|^2 \Big) \leq \\
\leq \sqrt{\frac{N}{D}}\epsilon_1 + \sum_j \Big(\epsilon + \|\Pi_N(X)|_{\mathcal{N}_D(p_{i_j})} - x_{i_j}\|^2 \Big) = \\
= \sqrt{\frac{N}{D}}(\epsilon_1 + \epsilon) + \|\Pi_N(X) - X\|^2.
\end{aligned}
\tag{24}
$$

Equation (24) proves the left hand side of (19) and this ends the proof. □

B. Implementation Details

In Section B.1 we describe more in detail the multiscale approach introduced in Section 5.1 of our submission. In Section B.2 we describe how we can avoid the computation of the nearest neighbors for many image patches. In this way we can decrease the computational complexity of our method.

B.1. Multiscale Approach

Given the score image X, the only parameter of our method is the size D of the patches x_i on which the projection is computed.

Ideally, we would like D to be large enough to capture enough contextual information. At the same time, using a too large value for D makes it difficult to gather a representative training set of patches. As a consequence, a large value of D can provoke loss of details. To handle this trade off, we adopted a multiscale approach. For clarity's sake, we describe below the case of 2 scales, but the generalization to an arbitrary number of scales is straightforward.

Given two patch sizes $D_1 > D_2 > 0$, for every pixel p_i we consider the patch $x_i = X(\mathcal{N}_{D_1}(p_i))$ of size D_1 and its central part of size D_2, $x_i^{(\text{cent})} = X(\mathcal{N}_{D_2}(p_i))$. Then, we consider the downsampled version of x_i to size D_2, $x_i^{(\text{down})}$.

We do this in both score images and for training ground truth patches $y \in \mathcal{M}_{D_1}$. We then perform Nearest Neighbors search in terms of the distance

$$k(x_i, y) = \|x_i^{(\text{cent})} - y^{(\text{cent})}\|^2 + \|x_i^{(\text{down})} - y^{(\text{down})}\|^2. \tag{25}$$

We then take the multiscale projection $\Pi_{D_1/D_2}(x_i)$ to be $y_{i_*}^{(\text{cent})}$, where $y_{i_*} = \arg\min k(x_i, y)$ in $\mathcal{M}_{D_1}$. This replaces the projection $\Pi_D(x_i)$ in the definition of $\Pi_{D\to N}(X)$.

Thanks to the downsampling term in Eq. (25) we can include more contextual information in the method. At the same time, by considering only the smaller central part $y_i^{(\text{cent})}$ for the final projection, we can preserve the details.

B.2. Efficient Implementation

Given the score map X, the main computational cost of our method comes from the computation of the nearest neighbors $\{\Pi_D(x_i)\}_{i=1}^N$, for every patch of size D in image X.

Many algorithms for approximated nearest neighbor search have been proposed [1, 2, 6] and can be applied in combination with our method. Since the size D used in our experiments can be very large, especially for 3D data, we use in our implementation the FLANN library [6], which is optimized for nearest neighbors search of high dimensional vectors.

Moreover, we take advantage of the specific properties of the ground truth manifold $\mathcal{M}_N$, and in particular of the sparsity of the ground truth images, to further reduce the computational cost. More precisely, suppose that the patch of zeros $\mathbf{0}$ of size D belongs to $\mathcal{M}_D$. Then, given an image patch x_i, it is easy to show that

$$\text{if} \quad \max_{p\in\mathcal{N}(p_i)} x_i(p) < \min_{y\in\mathcal{M}_D\setminus\{\mathbf{0}\}} \frac{\|y\|_2^2}{2\|y\|_1}, \quad \text{then} \quad \Pi_D(x_i) = \mathbf{0}. \tag{26}$$

This means that we do not need to explicitly do nearest neighbors search patches whose maximum is smaller than a given threshold, where the threshold can be computed in closed form from the training set $\mathcal{M}_D$.

To prove the statement above, we start by observing that

$$\Pi_D(x_i) = \mathbf{0} \Leftrightarrow 2 \sum_{p\in\mathcal{N}_D(p_i)} x_i(p)y(p) < \sum_{p\in\mathcal{N}_D(p_i)} y(p)^2, \ \forall y \in \mathcal{M}_D \setminus \{\mathbf{0}\}. \tag{27}$$

In fact, $\Pi_D(x_i) = \mathbf{0}$ if and only if $\|x_i - \mathbf{0}\|^2 < \|x_i - y\|^2$, for all $y \in \mathcal{M}_D \setminus \{\mathbf{0}\}$. Writing explicitely the distances, we have $\Pi_D(x_i) = \mathbf{0}$ if and only if

$$\sum_p x_i(p)^2 < \sum_p (x_i(p) - y(p))^2 = \sum_p \Big(x_i(p)^2 - 2x_i(p)y(p) + y(p)^2\Big). \tag{28}$$

Subtracting the left habd side from both sides of (29),

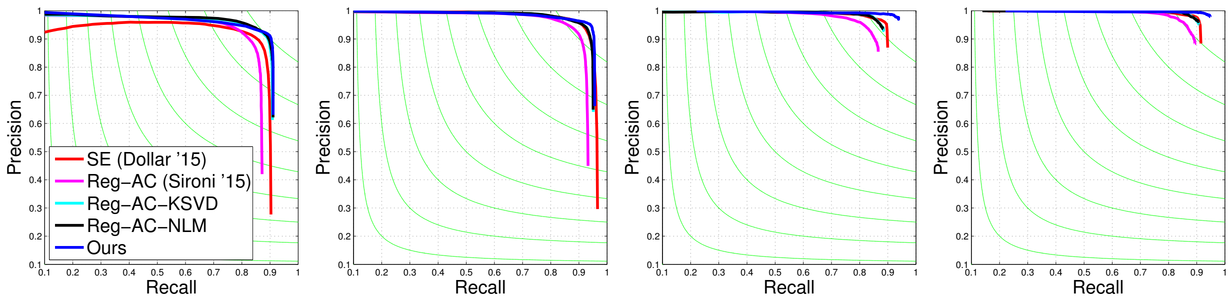
$$0 < -2\sum_p x_i(p)y(p) + \sum_p y(p)^2 \Leftrightarrow 2\sum_p x_i(p)y(p) < \sum_p y(p)^2. \tag{29}$$

This gives us condition (27).

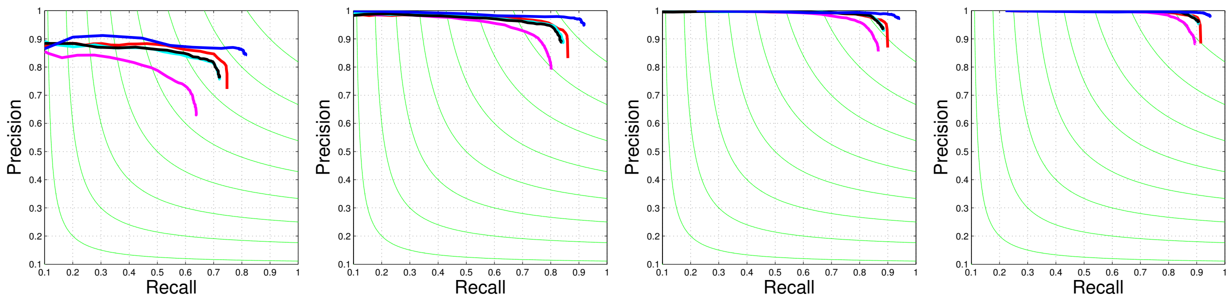
Now let $x_{\text{max}} = \max_p x_i(p)$. Since $y(p) \geq 0$ for every p[3],

$$\sum_p x_i(p)y(p) \leq \sum_p x_{\text{max}} y(p) = x_{\text{max}} \sum_p y(p). \tag{30}$$

[3]Form the definition of d in Eq. (1), we have $y(p) \geq 0$. For generic $y(p)$ we should consider $x_{\text{max}} = \max_p |x_i(p)|$



(a) tol = 1 (b) tol = 2 (c) tol = 3 (d) tol = 4



(a) tol = 1 (b) tol = 2 (c) tol = 3 (d) tol = 4

Figure 1. Precision-Recall curves for Centerline Detection on the Roads dataset for different tolerance values (in pixels). Top: evaluation on the whole image. Bottom: Evaluation on junctions only. Best viewed in color.

Thus, from (30) and (27)

$$\sum_p y(p)^2 > 2x_{\max} \sum_p y(p) \Rightarrow \sum_p y(p)^2 > 2 \sum_p x_i(p)y(p) \Rightarrow \Pi_D(x_i) = \mathbf{0}, \tag{31}$$

where the last implication follows by (27). Therefore, since $y(p) \geq 0$ and $y \neq \mathbf{0}$, we have $\sum_p y(p) = \|y\|_1 > 0$ and we can write (31) as

$$\text{if} \quad x_{\max} < \frac{\sum_p y(p)^2}{2\sum_p y(p)} \quad \forall y \in \mathcal{M}_D \setminus \{\mathbf{0}\}, \quad \text{then} \quad \Pi_D(x_i) = \mathbf{0}, \tag{32}$$

that is condition (26) we wanted to prove.

C. Additional Results

In this section we present additional results that, because of space limitations, did not fit in our submission.

Fig. 1 shows the Precision-Recall curves for the different methods on the centerline detection task of Section 5.1. The top row corresponds to the curves computed on the whole image for the different tolerance values considered. The bottom row shows analogous plots for the junctions evaluation framework. Fig. 2 shows the F-measure as a function of the tolerance factor and Fig. 3 the results on some test images.

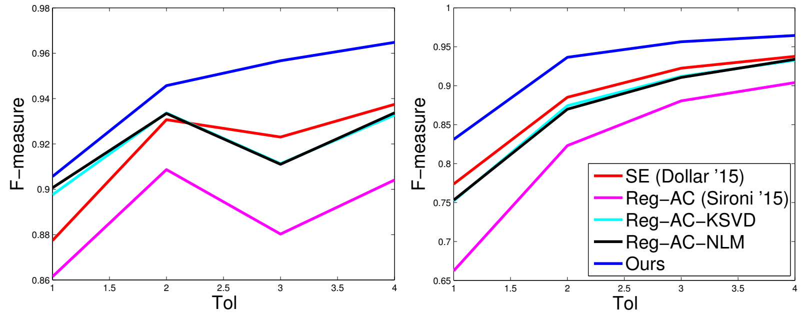
Fig. 4 shows the Precision-Recall curves for the different methods on the vessel segmentation task of Section 5.2 and Fig. 5 shows the results on some test images.

Fig. 6 shows the Precision-Recall curves for the different methods on the membrane detection task of Section 5.3 and the F-measure as a function of the tolerance factor. Fig. 7 shows the results on some test slices.

Finally, Fig. 8 shows the Precision-Recall curves for the different methods on the boundary detection task of Section 5.4 and Fig. 9 shows the results on some test images.

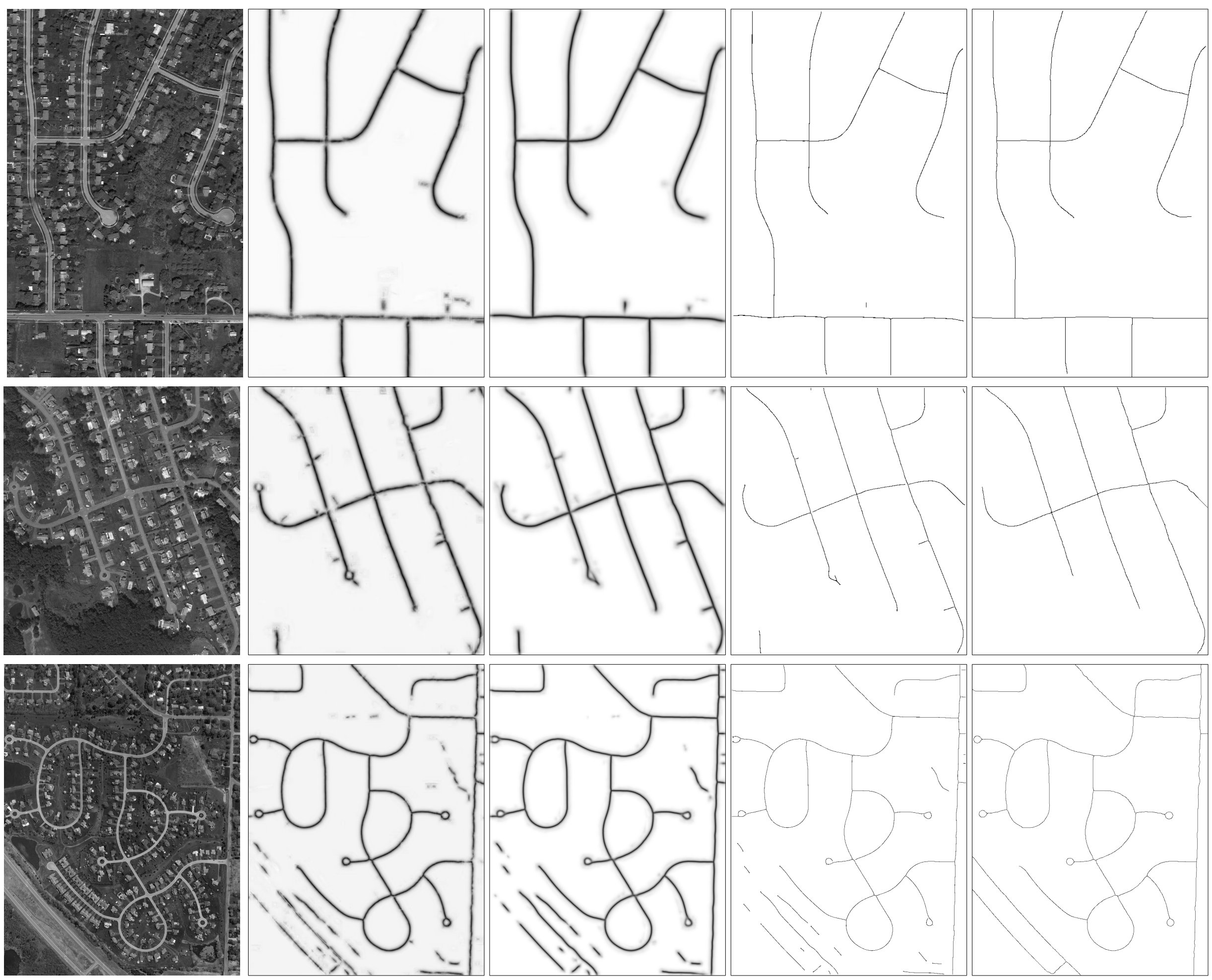
References

[1] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing. *ACM Trans. Graph.*, 2009.



(a) (b)

Figure 2. F-measure as a function of the tolerance factor for Centerline Detection on the Roads dataset. (a) Whole image; (b) Junctions only. Best viewed in color.



(a) (b) (c) (d) (e)

Figure 3. Centerline detection results. (a) Image; (b) **Reg-AC** [7] Score map; (c) **Ours** Score map; (d) Centerlines found after non-maxima suppression on (c) and thresholding; (e) Ground Truth.

[2] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The Generalized PatchMatch Correspondence Algorithm. In *ECCV*, 2010.

[3] C. Becker, K. Ali, G. Knott, and P. Fua. Learning Context Cues for Synapse Segmentation. *TMI*, 2013.

[4] P. Dollár and C. L. Zitnick. Fast Edge Detection Using Structured Forests. *PAMI*, 2015.

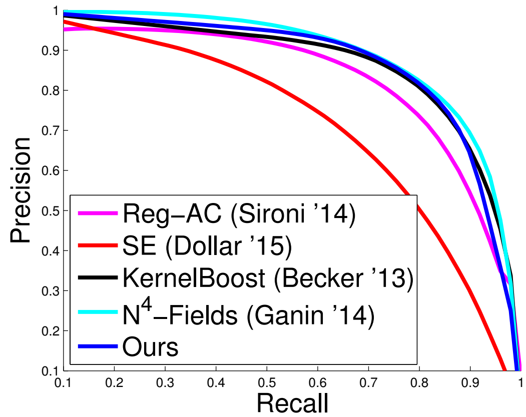
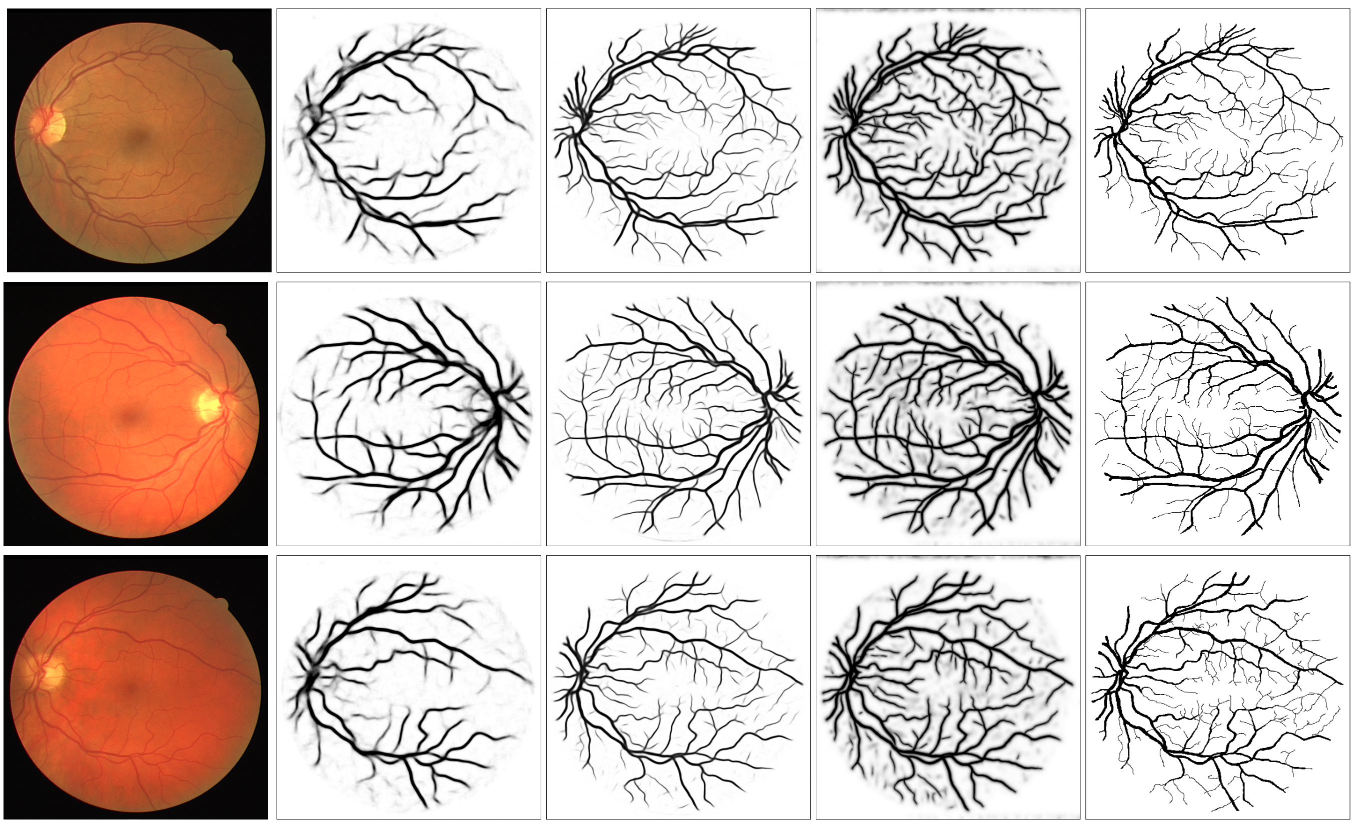


Figure 4. Precision-Recall curves for DRIVE dataset. Best viewed in color.



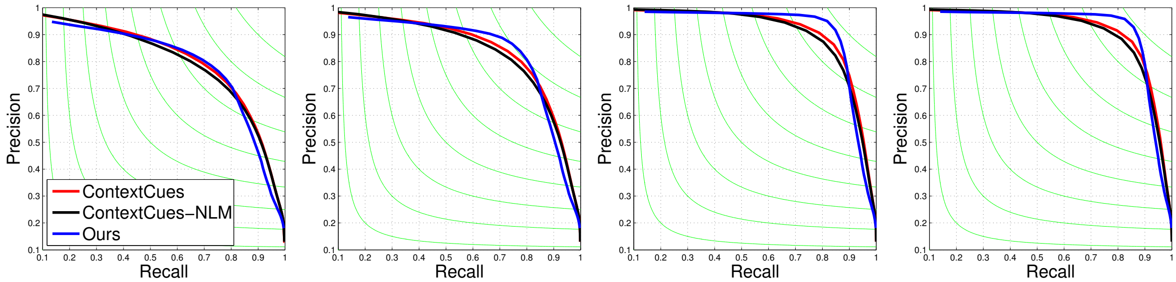
(a) (b) (c) (d) (e)

Figure 5. DRIVE dataset results. (a) Image; (b) **SE** [4]; (c) $\mathbf{N^4}$**-Fields** [5]; (d) **Ours**; (e) Ground Truth. Our method responds strongly on thin vessels.

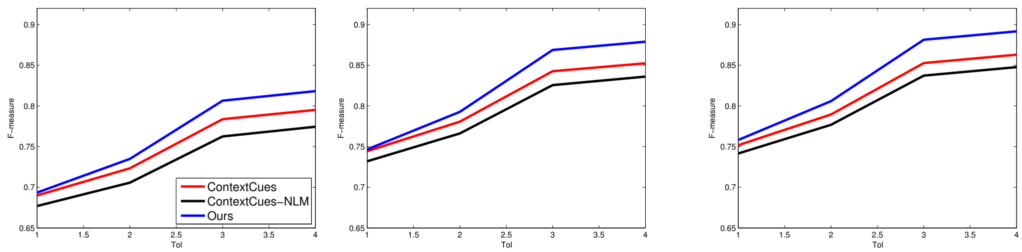
[5] Y. Ganin and V. Lempitsky. n^4-Fields: Neural Network Nearest Neighbor Fields for Image Transforms. In *ACCV*, 2014.

[6] M. Muja and D. G. Lowe. Scalable Nearest Neighbor Algorithms for High Dimensional Data. *PAMI*, 2014.

[7] A. Sironi, E. Turetken, V. Lepetit, and P. Fua. Multiscale Centerline Detection. *PAMI*, 2015.

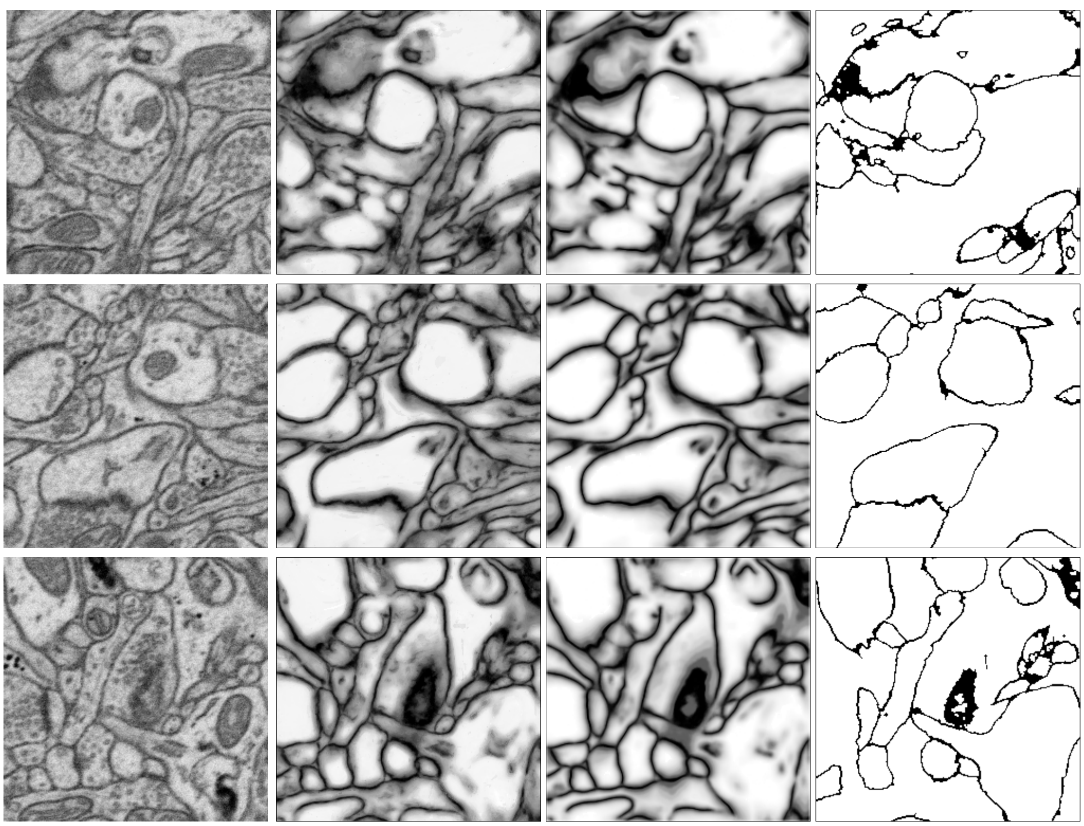


(a) tol = 1 (b) tol = 2 (c) tol = 3 (d) tol = 4



(a) No Auto-context (b) Auto-context iter 1 (c) Auto-context iter 2

Figure 6. Membrane Detection Results. Top: Precision-Recall curves at the second Auto-context iterations for different tolerance values (in voxels). Bottom: F-measure as a function of the tolerance factor for different Auto-context iterations. Best viewed in color.



(a) (b) (c) (d)

Figure 7. Membrane dataset results for three sample test slices. (a) Image; (b) **ContextCues** [3]; (c) **Ours**; (d) Ground Truth. Our method removes background noise while shapening the response on the membranes. Notice that ground truth volumes are only partially annotated.

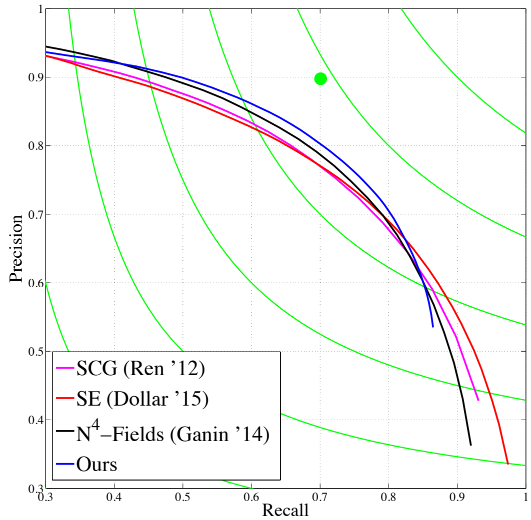


Figure 8. Precision-Recall curves for boundary detection on BSDS dataset. Best viewed in color.

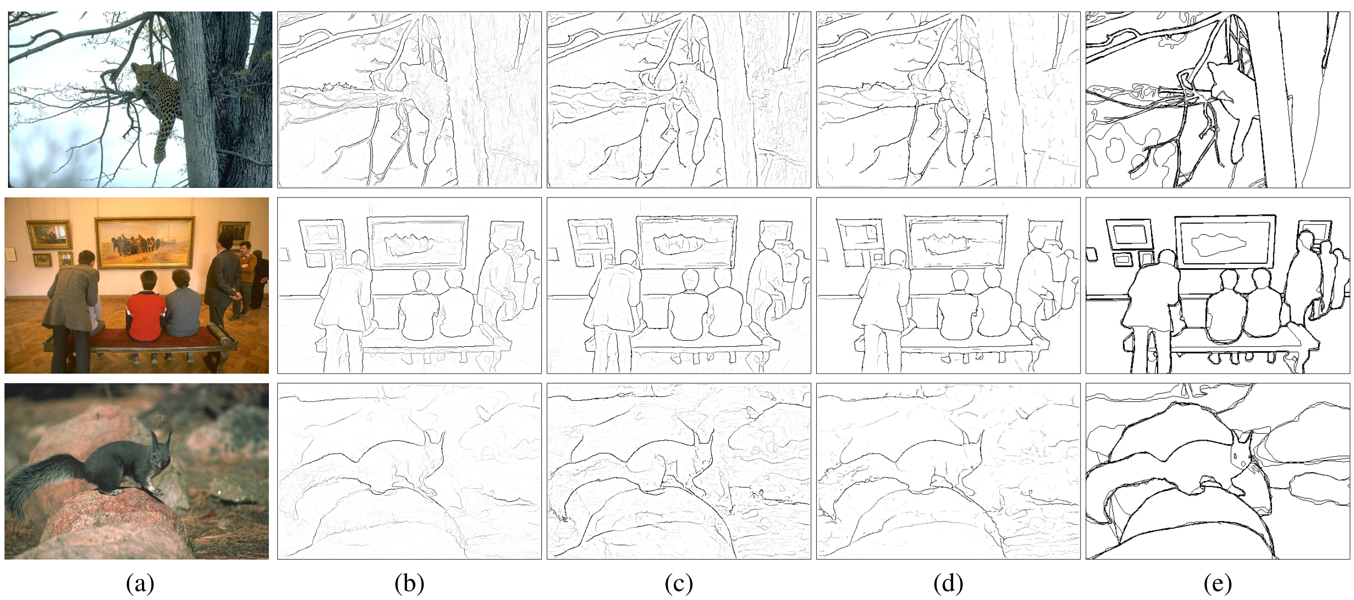


Figure 9. BSDS dataset results. (a) Image; (b) **SE** [4]; (c) **N^4-Fields** [5]; (d) **Ours**; (e) Human annotations. Our approach returns more continuous boundaries and preserves important details. Contrast has been enhanced for visualization purposes.