Assignment 4

Sagar Poudel - 213051001
Sejal Upadhye - 213050017

April 2022

1

1.1

Given equation 7 is:

$$J_3(v) = \sum_j \rho_j \left(A_{j\rightarrow} v - b_j\right)$$

where, $A_{j\rightarrow}$ is a matrix which has the rows that correspond to the derivative filters centered on pixel j that is applied to vectorized version of Image I_1 denoted as $\mathbf{v}$, and vector $\mathbf{b}$ denotes the input image derivative or zero.

In the above equation, when $j \in S_1$, we first need to subtract the gradient from the input image I and apply ρ. Similarly, when $j \in S_2$ which just need apply derivative filter to I_1. Thus when $j \in S_1$ we have derivative filter of I centered at j and when $j \in S_2$ the derivative filter of I_1 centered at pixel j. Also for the $j \in S_2$ the value of b become zero.

1.2

Equation 6 is given as:

$$\begin{aligned} J_2\left(I_1\right) = & \sum_{i,k} \rho\left(f_{i,k} \cdot I_1\right) + \rho\left(f_{i,k} \cdot \left(I - I_1\right)\right) \\ & + \lambda \sum_{i \in S_1, k} \rho\left(f_{i,k} \cdot I_1 - f_{i,k} \cdot I\right) \\ & + \lambda \sum_{i \in S_2, k} \rho\left(f_{i,k} \cdot I_1\right) \end{aligned}$$

In the above equation, the likelihood terms are:

$$\sum_{i,k} \rho\left(f_{i,k} \cdot I_1\right) + \rho\left(f_{i,k} \cdot \left(I - I_1\right)\right)$$

The paper uses the Laplacian mixture model as the likelihood filters.

The Prior terms are:

$$\sum_{i \in S_1, k} \rho\left(f_{i,k} \cdot I_1 - f_{i,k} \cdot I\right), \sum_{i \in S_2, k} \rho\left(f_{i,k} \cdot I_1\right)$$

The paper uses the log-histogram of derivative filters as the prior model.

1.3

The paper exploits the Statistical feature of the image saying that log-histogram is more sparse and compare to the Gaussian distribution. The given graph shows that log-histogram lies below the straight line connecting the minimal and maximal values. Here Laplacian lies at the border and Gaussian distribution lies above the straight line.

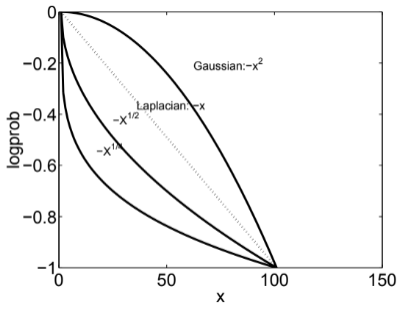

Figure 1: Sparsity with respect to different distributions. The Laplacian distribution is exactly at the middle between sparse and non-sparse. Gaussian lies above the Laplacian which is non sparse compare to log-histogram, which lies below the Laplacian which is more sparse.

Given in the paper that, distributions that are above the straight line will prefer to split an edge of unit contrast into two edges (one in each layer) with half the contrast, while distributions below the line will prefer decompositions in which the edge only appears in one of the layers but not in the other. We will refer to distributions that have this property in the log domain as being sparse.

2

The compressive measurement based on the Bayesian statistics is given by:

$y = \Phi x + \eta$

$y \in R^m, \Phi \in R^{m \times n}, \Phi \sim N\left(0, 1/m\right), x \in R^n, m << n$

$\eta \in R^m, \eta \sim N\left(0, \sigma^2 I_{m \times m}\right)$ where $\sigma = 0.01 *$ mean absolute value of Φx

X is random draw from a zero-mean Gaussian distribution with known co-variance matrix Σ, where $\mathbf{x} \in N(0, \Sigma)$

Considering the MAP solution for x given y and Φ:

$$\hat{\boldsymbol{x}} = argmax_{\boldsymbol{x}} p(x, \Phi | y)$$

Using the Bayes theorem, we get:

$$argmax_{\boldsymbol{x}} p(x, \Phi | y) = p(y | x, \Phi) p(x)$$

$$= argmax_{\boldsymbol{x}} \exp\left(-\|\boldsymbol{y} - \Phi \boldsymbol{x}\|^2 / 2\sigma^2\right) \frac{\exp\left(-\frac{\boldsymbol{x}^T \boldsymbol{\Sigma}_{\boldsymbol{x}}^{-1} \boldsymbol{x}}{2}\right)}{(2\pi)^{N/2} \left|\Sigma_x\right|^{0.5}}$$

Taking the log of the expression, due to negative term changes into the minimization problem

$$\boldsymbol{x} = argmax_{\boldsymbol{x}}(||\boldsymbol{y} - \boldsymbol{\Phi x}||^2 / 2\sigma^2 + \frac{1}{2} \boldsymbol{x}^T {\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1} \boldsymbol{x}) + constants$$

Taking the derivative wrt x and making it zero, we get,

$$2\boldsymbol{\Phi^T \Phi x} - 2\boldsymbol{y \Phi^T} / 2\sigma^2 + \frac{1}{2}(2 {\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1} \boldsymbol{x}) = 0$$

Solving the equation, we get,

$$(\boldsymbol{\Phi^T \Phi x} / 2\sigma^2 + {\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1} \boldsymbol{x}) = \boldsymbol{\Phi^T y} / 2\sigma^2$$

$$\hat{\boldsymbol{x}} = (\boldsymbol{\Phi^T \Phi} / 2\sigma^2 + {\boldsymbol{\Sigma}_{\boldsymbol{x}}}^{-1} / 2)^{-1} \boldsymbol{\Phi^T y} / (2\sigma^2)$$

In the plot, log(RMSE) was taken for the better representation. When $\alpha \geq 2$, the error was smaller.

Figure 2, provides the detail about RMSE with respect to alpha = 3 and 0, with the increase of the alpha value, the RMSE loss decreases Figure 3, provides the detail about the RMSE with respect to increase in α.

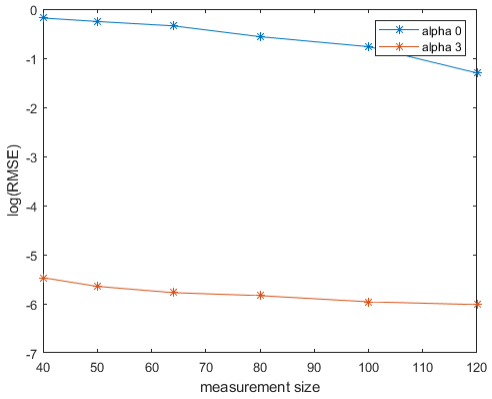

Figure 2: Plot of log(RMSE) vs # of measurement wrt alpha = 3 and 0

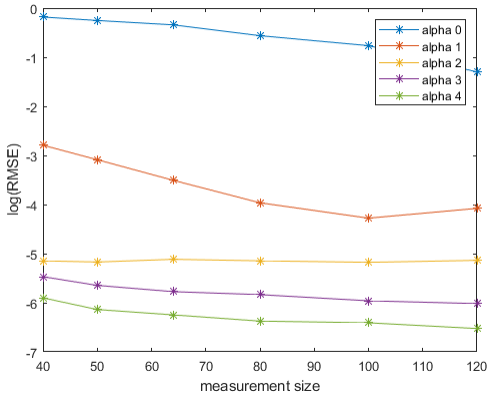

Figure 3: Plot of log(RMSE) vs # of measurement with increase value of alpha

We can observe that in both graph that, with the increase of α the RMSE decreases.

In our case, given $\sum_x = U\Lambda U^t$, and Λ is a diagonal matrix of eigenvalues of the form $\mathbf{ci}^{-\alpha}$. The eigenvalues of co-variance matrix $\sum_x$ decreases with increase of row index number which decays faster with increase of α. Due to this, the most of the values in $\mathbf{x}$ is close to zero or is zero. Thus making it easier to reconstruct x with increase of α. With increase of the measurement size, more number of observation is done, leading toward more accurate reconstruction. Thus RMSE reduces.
Thus due to above reason the RMSE is decreases with increase of α and measurement value.

3

3.1

Given, $A : \mathbf{R}^{m\times n} -> \mathbf{R}^p$ be the linear map and X_0 be the given matrix with rank r, and $b := A(X_0)$ and defines the convex problem,

$$X^* := \arg\min_X ||X||_* \quad \text{s.t. } A(X) = b$$

Here X^* is a matrix with minimum nuclear norm with subject to $A(X) = b$, while X_0 satisfy with $A(X_0) = b$. Thus $||X||_0$ should be greater than $||X^*||_0$

Therefore,

$$||X_0||_* \geq ||X^*||_*$$

3.2

From the theorem 3.3 we can see that, X_0 and R'_c are of same dimension and satisfy the constraint that $X_0 R'_C = 0$ and $X'_0 R_c = 0$. The given condition satisfies the Lemma 2.3. Thus we get,

$$||X_0 + R_c||_* = ||X_0||_* + ||R_c||_*$$

Hence, we get

$$\|X_0\|_* \geq \|X_0 + R\|_* \geq \|X_0 + R_c\|_* - \|R_0\|_* = \|X_0\|_* + \|R_c\|_* - \|R_0\|_*$$

Using the above equality we get,

$$\|X_0\|_* = \|X_0\|_* + \|R_c\|_* - \|R_0\|_*$$

Solving the equation we get,

$$\|R_0\|_* \geq \|R_c\|_*$$

3.3

In this problem, we assume when SVD of R_C is taken, where the singular values are arranged in a descending order along with diagonal. The left and right singular vectors are arranged accordingly. The partition of R_c is taken in such a way that the set is disjoint and have a rank atmost 3r.

Let σ_k be singular value of R_c and $k \in I_{i+1}$, similarly, σ_j be the singular value of R_c and $j \in I_i$. From the above assumption, we can write that:

$$\sigma_k \leq \sigma_j$$

Adding all the equivalent σ_j that may belong to I_i. we get $3r\sigma_k$ equals to summation. i,e

$$3r\sigma_k \leq \sum_{j \in I_i} \sigma_j$$

Thus,

$$\sigma_k \leq \frac{1}{3r} \sum_{j \in I_i} \sigma_j \forall\ k \in I_{i+1}$$

3.4

We know that, $||R_i||_* = \sum_{j \in I_i} \sigma_j$ as its the sum of singular value. Thus we get, using part 3 equation and squaring both terms

$$\sigma_k^2 \leq \frac{1}{9r^2} \|R_i\|_*^2 \ \forall k \in I_{i+1}$$

Now adding all the above 3r equations we get,

$$\sum_{k \in I_{i+1}} \sigma_k^2 \leq \frac{3r}{9r^2} ||R_i||_*^2$$

But, sum of square of the singular values gives the Frobenius, thus

$$\|R_{i+1}\|_F^2 \leq \frac{1}{3r} \|R_i\|_*^2$$

3.5

Using above part 5 inequality we can write by changing the variable,

$$||R_{j+1}||_F \leq \frac{1}{\sqrt{3r}} ||R_j||_*$$

Adding all the terms , we get

$$\sum_{j \geq 1} ||R_{j+1}||_F \leq \frac{1}{\sqrt{3r}} \sum_{j \geq 1} ||R_j||_*$$

We can write the above equation,

$$\sum_{j \geq 2} ||R_j||_F \leq \frac{1}{\sqrt{3r}} \sum_{j \geq 1} ||R_j||_*$$

3.6

As given $< R_i, R_j >= 0$ if $i \neq j$, thus using additive property of nuclear form on above part 5 equation we get,

$$\sum_{j\geq 1} ||R_j||_* = ||R_c||_*$$

Applying this equation on part 5 equation and substituting the value from part 2 final equation we get,

$$\sum_{j\geq 2} \|R_j\|_F \leq \frac{1}{\sqrt{3r}} \sum_{j\geq 1} \|R_j\|_* = \frac{1}{\sqrt{3r}} \|R_c\|_* \leq \frac{1}{\sqrt{3r}} \|R_0\|_*$$

3.7

We know that X is a r-rank matrix, thus using the condition 2 of Lemma 3.4 we can tell that, $rank(R_0) \leq 2r$ Using the relation of nuclear norm and Frobenius, we can write,

$$||R_0||_* \leq \sqrt{2r}\, ||R_0||_F$$

Using this equation in the above equation from the part 6 we get,

$$\sum_{j\geq 2} \|R_j\|_F \leq \frac{1}{\sqrt{3r}} \sum_{j\geq 1} \|R_j\|_* = \frac{1}{\sqrt{3r}} \|R_c\|_* \leq \frac{1}{\sqrt{3r}} \|R_0\|_* \leq \frac{\sqrt{2r}}{\sqrt{3r}} \|R_0\|_F$$

3.8

Given in the paper, we know that, $rank(R_0) \leq 2r$ and $R1, R2, ...,$ each of rank at most 3r. we can write

$$rank(R_0 + R_1) \leq 2r + 3r = 5r$$

3.9

we know that $R = R_0 + R_c$ We also know that

$$R_c = R_1 + \sum_{j\geq 2} R_j$$

Then with triangle inequality,

$$\|\mathcal{A}(R)\| = \left\| \mathcal{A}\left((R_0 + R_1) + \sum_{j\geq 2} R_j \right) \right\| \geq \|\mathcal{A}(R_0 + R_1)\| - \left\| \mathcal{A}\left(\sum_{j\geq 2} R_j \right) \right\|$$

Using the same triangle inequality to last term we get,

$$\left\| \mathcal{A}\left(\sum_{j\geq 2} R_j \right) \right\| = \mathcal{A}\left(\left\| \sum_{j\geq 2} R_j \right\| \right)$$

Thus we get,

$$\|\mathcal{A}(R)\| = \left\| \mathcal{A}\left((R_0 + R_1) + \sum_{j\geq 2} R_j \right) \right\| \geq \|\mathcal{A}(R_0 + R_1)\| - \mathcal{A}\left(\left\| \sum_{j\geq 2} R_j \right\| \right)$$

3.10

As using the rank given in part 8 we can write,

$$\|\mathcal{A}(R_0 + R_1)\| \geq (1 - \delta_{5r}) \|R_0 + R_1\|_F$$

and

$$\|\mathcal{A}(R_j)\| \leq (1 + \delta_{3r}) \|R_j\|_F \quad \forall j \geq 2$$

Applying the above relation to part 9 we get,

$$||A(R)|| \geq (1 - \delta_{5r}) \|R_0 + R_1\|_F - (1 + \delta_{3r}) \sum_{j\geq 2} \|R_j\|_F$$

3.11

We know that $R = X^* - X_0$ and for A we have $A(X*) = b$ and $A(X_0) = b$ Now solving we get,

$$A(R) = A(X^* - X_0) = A(X^*) - A(X_0) = b - b = 0$$

3.12

We have $A(R) = 0$

$$||A(R)|| \geq \left((1 - \delta_{5r}) - \frac{9}{11}(1 + \delta_{3r})\right) \|R_0\|_F$$

We know Frobenius norm is positive, to make right-hand side positive we must have following term positive,

$$\left((1 - \delta_{5r}) - \frac{9}{11}(1 + \delta_{3r})\right) > 0$$

Solving the equation we get,

$$9\delta_{3r} + 11\delta_{5r} < 2.$$

4

4.1

To recover a low-rank matrix, the matrix cannot be in the null space of the sampling operator giving the value of subset of the entries. If the singular vectors of a matrix M are highly concentrated. Considering the rank-2 symmetric matrix given by:

$$\boldsymbol{M} = \sum_{k=1}^{2} \sigma_k \boldsymbol{u}_k \boldsymbol{u}_k^*, \quad \begin{array}{l} \boldsymbol{u}_1 = (\boldsymbol{e}_1 + \boldsymbol{e}_2)/\sqrt{2} \\ \boldsymbol{u}_2 = (\boldsymbol{e}_1 - \boldsymbol{e}_2)/\sqrt{2} \end{array}$$

where the singular values are arbitrary, this matrix vanishes excepts in the top-left corner and one would basically need to see all the entries of **M** to be able to recover this matrix exactly.

Thus the paper considers low coherence, as matrices whose row and column have low coherence cannot really be in the null space of the sampling operator.

4.2

When the main result extends to a variety of other low rank matrix completion problems beyond the sampling of entries, we get new constraint rank minimization problem, given by:

$$\begin{array}{ll} \text{minimize} & \text{rank}(\boldsymbol{X}) \\ \text{subject to} & \boldsymbol{f}_i^* \boldsymbol{X} \boldsymbol{g}_j = \boldsymbol{f}_i^* \boldsymbol{M} \boldsymbol{g}_j, \quad (i,j) \in \Omega \end{array}$$

Suppose M is of rank r, we can decompose M and get,

$$\boldsymbol{M} = \sum_{k=1}^{r} \sigma_k \boldsymbol{f}_i^* \boldsymbol{u}_k \boldsymbol{u}_k^* \boldsymbol{g}_j$$

In order to get the result, the row and the column space of the **M** must the respectively incoherent with the basis ($\boldsymbol{f_i}$) and $\boldsymbol{g_j}$ respectively.

4.3

The paper gives the example of rank-1 matrix **M** to show that one cannot hope to be able to recover a low rank matrix from a sample of its entries.
The matrix **M** is of rank M equal to

$$\boldsymbol{M} = \boldsymbol{e}_1 \boldsymbol{e}_n^* = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

where, e_i is the i_{th} canonical basis vector in Euclidean space. Matrix has 1 only at the top-right corner, rest all are 0s. This matrix cannot be recovered from a sampling of its entries unless we see all the entries because for most sampling set, we get to see zeros there is no way of guessing the matrix is not zero. If we see the 90% of entries randomly, we wish to see only 10% of the time zero. It is therefore impossible to recover all low-rank matrices from a set of sampled entries.

5

Title - A Unified Approach to Salient Object Detection via Low Rank Matrix Recovery
Venue - 2012 IEEE Conference on Computer Vision and Pattern Recognition

Salient object detection is not a pure low-level, bottom-up process. Higher-level knowledge is important even for task-independent image saliency. The methods like self-information, graphic models, log-spectrum and sparsity models may work well for low-level saliency (or saliency regions), but they are neither sufficient nor necessary, especially in the cases when the saliency is also related to the human perception or is task-dependent. While the salient regions are mostly unique, the inverse might not necessarily be true. Not all unique regions are salient, and a small region with high local contrast might be considered as meaningless noise by the human.

The paper proposes a unified model to incorporate traditional low-level features with higher-level guidance to detect salient objects. In the model, an image is represented as a low-rank matrix plus sparse noises in a certain feature space, where the non-salient regions (or background) can be explained by the low-rank matrix, and the salient regions are indicated by the sparse noises. To ensure the validity of this model, a linear transform for the feature space is introduced and needs to be learned. Given an image, its low-level saliency is then extracted by identifying those sparse noises when recovering the low-rank matrix. Furthermore, higher-level knowledge is fused to compose a prior map, and is treated as a prior term in the objective function to improve the performance. Extensive experiments show that the model can comfortably achieves comparable performance to the existing methods even without the help from high-level knowledge. The integration of top-down priors further improves the performance and achieves the state-of-the-art. Moreover, the proposed model can be considered as a prototype framework not only for general salient object detection, but also for potential task-dependent saliency applications.

- **Low-level Saliency Detection** Given an image, different types of visual features such as Color, Steerable pyramids Gabor filters around each pixel are extracted. All these features are then stacked vertically to form a feature vector. Then image segmentation is performed based on the extracted features by mean-shift clustering. The image is considered as a combination of a background residing in a low dimensional space with salient objects as sparse noises. Therefore, the feature matrix representation F can be decomposed into two parts $F = L + S$, where L is the low-rank matrix corresponding to the background while S is a s-parse matrix representing the salient regions. The low-rank matrix recovery problem can then be formulated as:

$$(\mathbf{L}^*, \mathbf{S}^*) = arg\min_{\mathbf{L},\mathbf{S}} (rank(\mathbf{L}) + \lambda \parallel \mathbf{S} \parallel_0) \quad s.t. \quad \mathbf{F} = \mathbf{L} + \mathbf{S} \tag{1}$$

Since the above problem is NP-hard and hard to approximate, one can alternatively solve the convex surrogate:

$$(\mathbf{L}^*, \mathbf{S}^*) = arg\min_{\mathbf{L},\mathbf{S}} (\parallel \mathbf{L} \parallel_* + \lambda \parallel \mathbf{S} \parallel_1) \quad s.t. \quad \mathbf{F} = \mathbf{L} + \mathbf{S} \tag{2}$$

where $||L||$ is the nuclear norm of L and $||.||_1$ indicates l_1-norm. L and S can be perfectly recovered by Eqn.2 in most cases by RPCA.The l_1-norm of each column S_i in S is used to measure the saliency of corresponding segments. A saliency map is then accordingly generated and normalized to be a gray-scale image. A linear transformation is then learnt on the original feature space T from a set of training images.

After transformation, features can be represented as $g_i = Tf_i$ where $T \in R^{D \times D}$. Accordingly, G = [g_1, g_2, ..., g_N] = TF, and the formulation in Eqn. 2 is advanced to:

$$(\mathbf{L}^*, \mathbf{S}^*) = arg \min_{\mathbf{L},\mathbf{S}} (\| \mathbf{L} \|_* + \lambda \| \mathbf{S} \|_1) \quad s.t. \quad \mathbf{TF} = \mathbf{L} + \mathbf{S} \tag{3}$$

It is infeasible to specify T, but it can be learned.

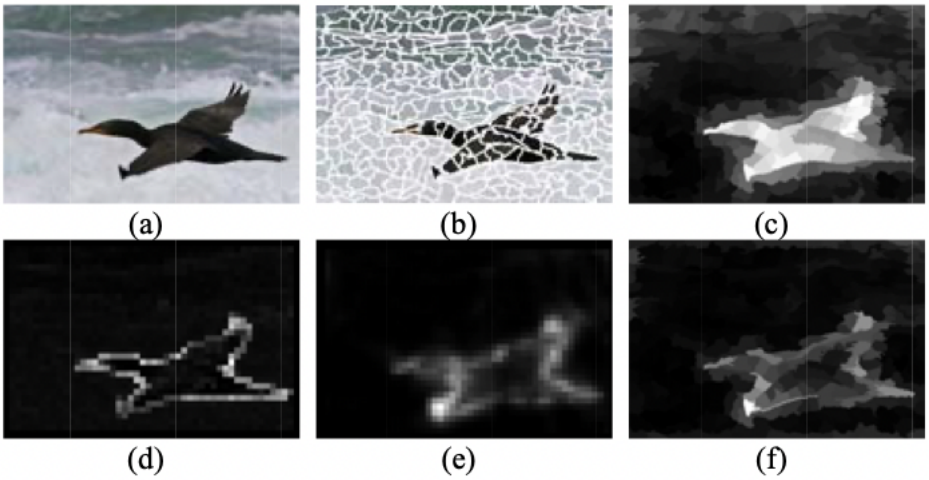

Figure 4: Illustration on low-level saliency detection by the model. (a) the original image, (b) over-segmentation result, (c) detected saliency with learned feature transformation, which is better than others, (d) detected saliency by SC, which only has high values on edges, (e) saliency by uniform sampling, which also cannot detect the entire object, (f) saliency by over-segmentation without feature learning, better than (d) and (e), but still has low saliency values inside the object.

- **Higher-level Prior Integration** The higher-levels are generally based on human perception. The following higher-level priors are generated and integrated to the model: Location prior, Semantic prior and Color prior. These prior maps are then multiplied together to produce a final prior map. The below figure gives an example of those individual prior maps and the fused high-level prior map.

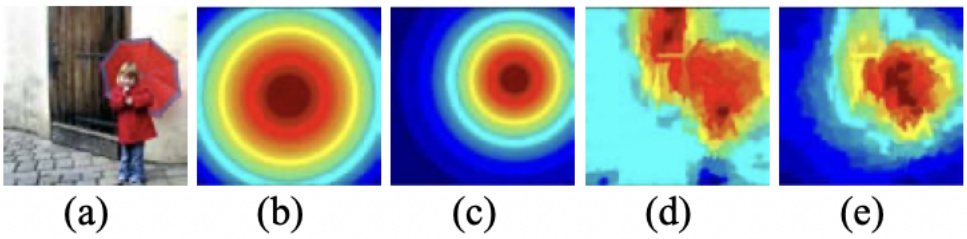

Figure 5: Example of high-level prior maps. (a) original image, (b) location prior map, (c) prior maps generated by face detection, (d) color prior map, (e) final fused prior map.

According to the prior map, we know the probability of being salient for each segment based on the location of the segment center, which is denoted by p_i. Such prior probability can also be represented by a diagonal matrix P = diag(p_1, p_2, ..., p_N). P can be naturally incorporated in the formulation:

$$(\mathbf{L}^*, \mathbf{S}^*) = arg \min_{\mathbf{L},\mathbf{S}} (\| \mathbf{L} \|_* + \lambda \| \mathbf{S} \|_1) \quad s.t. \quad \mathbf{TFP} = \mathbf{L} + \mathbf{S}$$

By integrating the high-level priors to the formulation, the effects to the saliency detection are two-fold:

– In P, most of p_i are relatively small. Therefore feature vectors multiplied by a larger p_i will be considered as outliers in the low rank matrix L and more likely to be included in the sparse noise matrix S.

– The error terms in S is also magnified with a larger p_i. Since our saliency map is generated according to the errors in S, regions with larger p_i tend to produce higher saliency.

Therefore regions with larger priors will be highlighted in the final saliency map by solving the above equation. (d) in below figure shows some results after integrating the high-level prior, which are better than the results without such information in (c).

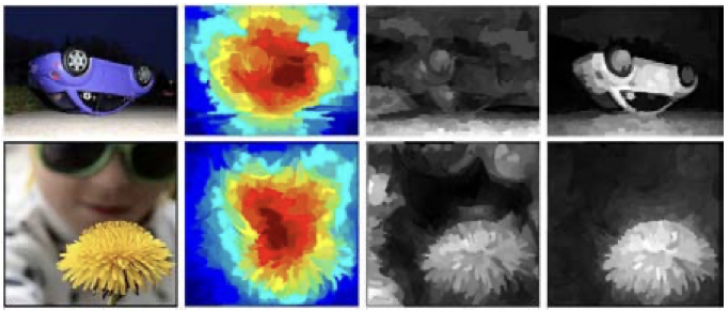

Figure 6: The integration of higher-level priors further improves the performance.

Moreover, by unifying low-level information in F and higher-level priors in P, the model is robust to incorrect top-down higher-level guidance to some extent. Consider the case when some parts of a homogenous background in an image are falsely assigned with large priors due to incorrect guidance from higher-level (see above figure for example, some background regions are also marked with high priors), corresponding feature vectors from these areas are still highly correlated to other background regions. As a result, they are not considered to be noises and will not be labeled as salient regions, as we can observe in above figure (d).