SPIKE-AND-SLAB VARIATIONAL INFERENCE FOR BLIND IMAGE DECONVOLUTION

Juan G. Serra[1], Javier Mateos[1], Rafael Molina[1], Aggelos K. Katsaggelos[2]

Contact email: jmd@decsai.ugr.es

[1]Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad de Granada, Spain.

[2]Electrical Engineering and Computer Science Department. Northwestern University, Evanston, IL, USA.

UNIVERSIDAD DE GRANADA

Northwestern University

1. Introduction

Blind image deconvolution (BID) aims at retrieving the original sharp image **x** from a blurry and noisy observation **y**.

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$$

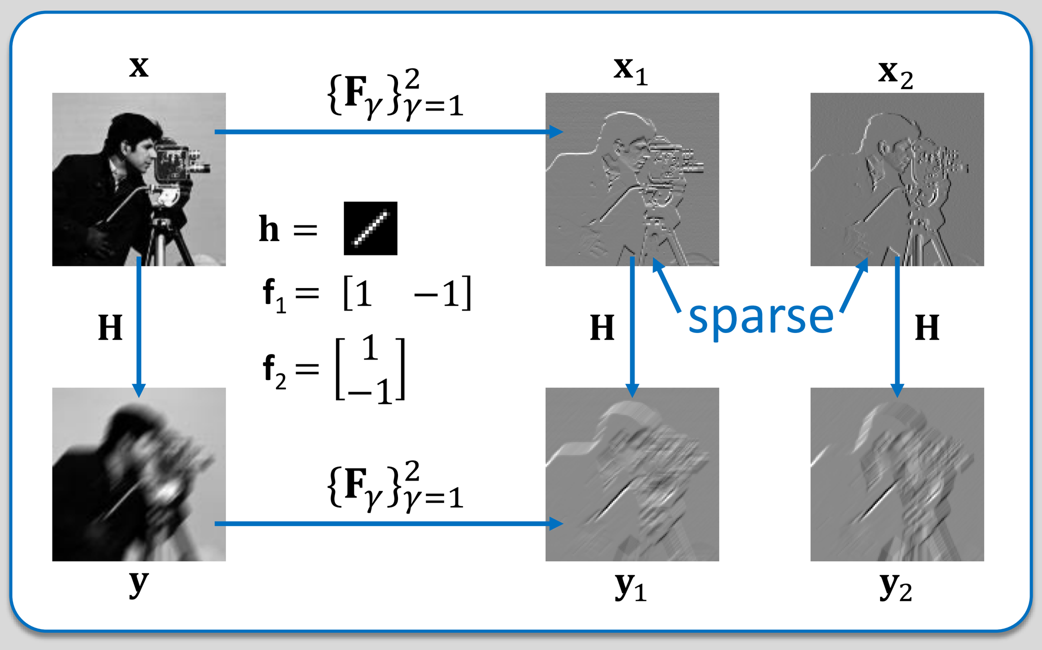
noise

where **x** denotes the original image and **H** is the convolution matrix associated with the unknown blur kernel **h**.

We formulate the BID problem in the filter space, creating a set of L pseudo-observations

$$\mathbf{y}_\gamma = \mathbf{F}_\gamma \mathbf{y} = \mathbf{H}\mathbf{F}_\gamma \mathbf{x} + \mathbf{F}_\gamma \mathbf{n} = \mathbf{H}\mathbf{x}_\gamma + \mathbf{n}_\gamma$$

Notice that $\{\mathbf{x}_\gamma\}_{\gamma=1}^{L}$ should be sparse since they represent high-pass filtered instances of the original image.



We develop a **Blind Image Deconvolution** method that
- uses **Bayesian Modeling and Variational Inference**
- utilizes the **Spike-and-Slab prior** to impose sparsity
- is more robust against noise

The variational inference we propose
- is **more accurate** than the standard mean field variational approximation
- is much **more efficient** than MCMC

2. Bayesian Modeling

Observation model (in the filter space)

noise precision

$$\mathrm{p}(\mathbf{y}_\Gamma|\mathbf{h},\mathbf{x}_\Gamma,\boldsymbol{\beta}_\Gamma) = \prod_{\gamma=1}^{L} \mathrm{p}(\mathbf{y}_\gamma|\mathbf{h},\mathbf{x}_\gamma,\beta_\gamma) = \prod_{\gamma=1}^{L} \mathcal{N}(\mathbf{y}_\gamma|\mathbf{H}\mathbf{x}_\gamma,\beta_\gamma^{-1}\mathbf{I})$$

Prior model

We impose a **Spike-and-Slab prior** on each pixel $x_{\gamma i}$

$$p(x_{\gamma i}|\alpha_\gamma,\pi_\gamma) = \pi_\gamma \mathcal{N}(x_{\gamma i}|0,\alpha_\gamma^{-1}) + (1-\pi_\gamma)\delta(x_{\gamma i}),$$

Slab: represents the edges of the image

Spike: Dirac delta, represents the flat areas of the image

This is a truly sparse prior: $x_{\gamma i}$ is exactly 0 with probability $1-\pi_\gamma$.

Unfortunately, variational inference with this prior is intractable.

However, we then rewrite $x_{\gamma i}$ as the product of a Gaussian zero-mean random variable $\tilde{x}_{\gamma i}$ and a Bernoulli random variable $s_{\gamma i}$.

$$x_{\gamma i} = s_{\gamma i}\tilde{x}_{\gamma i}$$

and redefine the prior on the two components of $x_{\gamma i}$ as

$$\mathrm{p}(\tilde{x}_{\gamma i}, s_{\gamma i}|\alpha_\gamma,\pi_\gamma) = \mathcal{N}(\tilde{x}_{\gamma i}|0,\alpha_\gamma^{-1})\pi_\gamma^{s_{\gamma i}}(1-\pi_\gamma)^{1-s_{\gamma i}}$$

where $s_{\gamma i} \in \{0,1\}$, which is tractable.

Joint probability distribution

With all the above, we have

Observation model

$$\mathrm{p}(\boldsymbol{\Theta},\mathbf{y}_\Gamma) = \mathrm{p}(\mathbf{h})\prod_\gamma \mathrm{p}(\mathbf{y}_\gamma|\mathbf{h},\tilde{\mathbf{x}}_\gamma,\mathbf{s}_\gamma,\beta_\gamma) \times \prod_\gamma \prod_i \mathrm{p}(\tilde{x}_{\gamma i}, s_{\gamma i}|\alpha_\gamma,\pi_\gamma).$$

set of unknowns $\Theta = \{\mathbf{h},\tilde{\mathbf{x}}_\Gamma,\mathbf{s}_\Gamma\}$

Spike-and-slab-prior

Flat prior on **h**

3. Variational Bayesian Inference

Since $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y})$ cannot be calculated in closed form, the standard mean field approximation that factorizes $\mathrm{q}(\tilde{\mathbf{x}}_\gamma,\mathbf{s}_\gamma) = \mathrm{q}(\tilde{\mathbf{x}}_\gamma)\mathrm{q}(\mathbf{s}_\gamma)$ could be used. However this is a unimodal distribution [1] and, therefore, not a good approximation of the true posterior distribution. Since the pairs $\{\tilde{\mathbf{x}}_\gamma,\mathbf{s}_\gamma\}$ are strongly correlated (recall that $x_{\gamma i} = \tilde{x}_{\gamma i}s_{\gamma i}$), we treat them as a unit, hence we use the factorization

$$\mathrm{q}(\mathbf{x}_\gamma,\mathbf{s}_\gamma) = \prod_{\gamma=1}^{L}\prod_{i=1}^{N}\mathrm{q}(\tilde{x}_{\gamma i}, s_{\gamma i})$$

and utilize the following mean field approximation

$$\mathrm{q}(\boldsymbol{\Theta}) = \mathrm{q}(\mathbf{h})\prod_{\gamma=1}^{L}\prod_{i=1}^{N}\mathrm{q}(\tilde{x}_{\gamma i}, s_{\gamma i}).$$

Obtaining $\mathrm{q}(\tilde{\mathbf{x}}_\gamma,\mathbf{s}_\gamma)$

Using the Kullback-Leibler criterion and the mean field factorization presented above, we have

$$\mathrm{q}(\tilde{x}_{\gamma i},s_{\gamma i}) = \frac{1}{\mathcal{Z}}\exp\left[\langle \ln \mathrm{p}(\mathbf{y}_\gamma|\mathbf{h},\tilde{\mathbf{x}}_\gamma,\mathbf{s}_\gamma,\beta_\gamma)\rangle_{\mathrm{q}(\boldsymbol{\Theta}_{\tilde{x}_{\gamma i},s_{\gamma i}})}\right] \times \mathcal{N}(\tilde{x}_{\gamma i}|0,\alpha_\gamma^{-1})\pi_\gamma^{s_{\gamma i}}(1-\pi_\gamma)^{1-s_{\gamma i}}.$$

To compute the explicit expression for the above posterior we separate the derivations for $\mathrm{q}(s_{\gamma i})$ and $\mathrm{q}(\tilde{x}_{\gamma i}|s_{\gamma i})$.

The distributions $\mathrm{q}(s_{\gamma i}=0)$ and $\mathrm{q}(s_{\gamma i}=1)$ can be obtained by marginalization. Defining

$$\omega_{\gamma i} = q(s_{\gamma i}=1) = \frac{1}{1+e^{-u_{\gamma i}}}$$

we have

$$u_{\gamma i} = \ln \mathrm{q}(s_{\gamma i}=1) - \ln \mathrm{q}(s_{\gamma i}=0) = \ln\frac{\pi_\gamma}{1-\pi_\gamma} + \frac{1}{2}\ln\alpha_\gamma - \frac{1}{2}\ln(\rho_\gamma) + \frac{\beta_\gamma^2}{2\rho_\gamma}\Big(\mathbf{h}_i^{\mathrm{T}}(\mathbf{y}_\gamma - \sum_{k\neq i}\langle s_{\gamma k}\tilde{x}_{\gamma k}\rangle\,\mathbf{h}_k)\Big)^2.$$

It can be shown that the conditional distributions $\mathrm{q}(\tilde{x}_{\gamma i}|s_{\gamma i})$ for $s_{\gamma i} = \{0,1\}$ are both Gaussians of the form

$$\mathrm{q}(\tilde{x}_{\gamma i}|s_{\gamma i}=0) = \mathcal{N}(\tilde{x}_{\gamma i}|0,\alpha_\gamma^{-1}),$$
$$\mathrm{q}(\tilde{x}_{\gamma i}|s_{\gamma i}=1) = \mathcal{N}(\tilde{x}_{\gamma i}|\mu_{x_{\gamma i}},\rho_\gamma^{-1}),$$

with

$$\mu_{x_{\gamma i}} = \frac{\beta_\gamma}{\rho_\gamma}\mathbf{h}_i^{\mathrm{T}}(\mathbf{y}_\gamma - \textstyle\sum_{k\neq i}\langle s_{\gamma k}\tilde{x}_{\gamma k}\rangle\,\mathbf{h}_k)$$
$$\rho_\gamma = \beta_\gamma \parallel \mathbf{h} \parallel^2 + \alpha_\gamma.$$

Finally, we can express the posterior as

$$\mathrm{q}(\tilde{x}_{\gamma i}, s_{\gamma i}) = \mathrm{q}(\tilde{x}_{\gamma i}|s_{\gamma i})\mathrm{q}(s_{\gamma i}) = \omega_{\gamma i}^{s_{\gamma i}}(1-\omega_{\gamma i})^{1-s_{\gamma i}} \times \mathcal{N}(\tilde{x}_{\gamma i} \mid s_{\gamma i}\mu_{x_{\gamma i}},\ s_{\gamma i}\rho_\gamma^{-1} + (1-s_{\gamma i})\alpha_\gamma^{-1}).$$

Furthermore,

$$\langle s_{\gamma i}\tilde{x}_{\gamma i}\rangle = \langle x_{\gamma i}\rangle = \omega_{\gamma i}\mu_{x_{\gamma i}},$$
$$\langle s_{\gamma i}^2\tilde{x}_{\gamma i}^2\rangle = \langle s_{\gamma i}\tilde{x}_{\gamma i}^2\rangle = \langle x_{\gamma i}^2\rangle = \omega_{\gamma i}(\mu_{x_{\gamma i}}^2 + \rho_\gamma^{-1}).$$

Obtaining $\mathrm{q}(\mathbf{h})$

Notice that we assume a degenerate distribution on $\mathrm{q}(\mathbf{h})$, which leads to the point estimate for **h**

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \sum_{\gamma=1}^{L}\langle ||\mathbf{y}_\gamma - \mathbf{H}\mathbf{x}_\gamma||^2\rangle_{\mathrm{q}(\boldsymbol{\Omega}_\mathbf{h})} = \arg\min_{\mathbf{h}}\sum_{\gamma=1}^{L}\Big[||\mathbf{y}_\gamma - \mathbf{H}\langle\mathbf{x}_\gamma\rangle||^2 + \sum_{j=1}^{N}(\langle x_{\gamma j}^2\rangle - \langle x_{\gamma j}\rangle^2)||\mathbf{h}||^2\Big],$$

constrained to $h_j \geq 0$, $\sum_j h_j = 1$.

We can efficiently solve this minimization problem with the ADMM method in [2].

Final image estimation

Once the estimate of the blur, $\hat{\mathbf{h}}$, has been obtained, a non-blind deconvolution algorithm is used to recover an estimation of the original sharp image by solving the problem

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\parallel \hat{\mathbf{H}}\mathbf{x} - \mathbf{y}\parallel_2^2 + \frac{\lambda}{p}\textstyle\sum_\gamma \parallel \mathbf{x}_\gamma \parallel_p,$$

using the fast iterative method in [2].

[1] M. K. Titsias and M. Lázaro-Gredilla, "Spike and slab variational inference for multi-task and multiple kernel learning," in NIPS, 2011, pp. 2339–2347.

[2] Zhou, M. Vega, F. Zhou, R. Molina, and A. K. Katsaggelos, "Fast Bayesian blind deconvolution with Huber super Gaussian priors," Digit. Signal Process., vol. 60, pp. 122–133, 2017.

4. Blind Deconvolution Algorithm

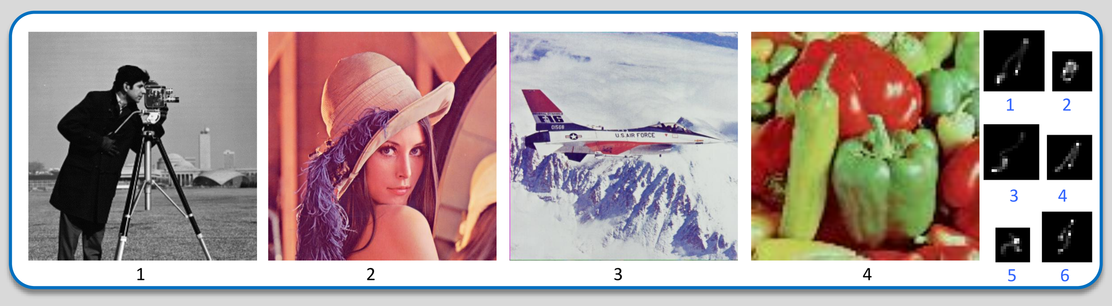
Algorithm 1 Blur estimation algorithm

Require: $\mathbf{y}_\Gamma$, $\boldsymbol{\beta}_\Gamma$, $\boldsymbol{\alpha}_\Gamma$, $\boldsymbol{\pi}_\Gamma$, and initial values for h^0, and $\langle\mathbf{x}_\Gamma\rangle^{(0)}$.
Set $k = 0$.
repeat
  Update $\mu_{x_{\gamma i}}^{(k)}$, $\rho_\gamma^{(k)}$, and $\omega_{\gamma i}^{(k)}$, $\forall\gamma$, $\forall i$.
  Update $\langle x_{\gamma i}\rangle$ and $\langle x_{\gamma i}^2\rangle$, $\forall\gamma$, $\forall i$.
  Update $\mathrm{h}^{(k+1)}$.
  Set $k = k+1$.
until convergence.
return $\hat{\mathrm{h}} = \mathrm{h}^{(k)}$.

We embedded the blur estimation into a multiscale scheme.
Once we estimate $\hat{\mathbf{h}}$, we use a non-blind deconvolution algorithm to estimate the original sharp image.

5. Experimental Results

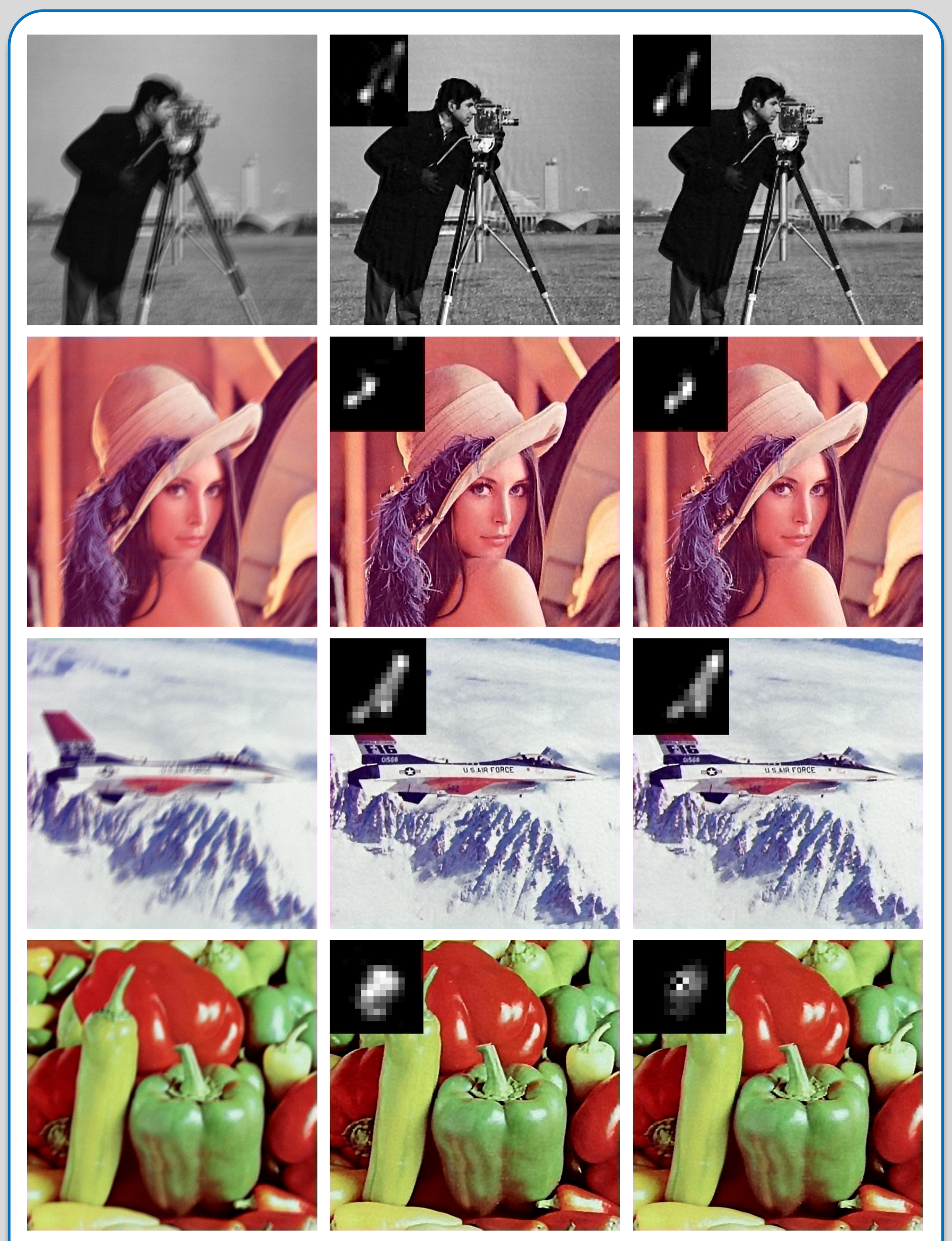
Test were run on a set of 4 test images with the 6 blur kernels.



$\boldsymbol{\beta}_\Gamma = 5000$, its real value, $\boldsymbol{\alpha}_\Gamma, \boldsymbol{\pi}_\Gamma$ were selected by a grid search. We used 2 high-pass filters: $f_1 = [1,-1]$ and $f_2 = [1,-1]^{\mathrm{T}}$ for the blur estimation. For the final image reconstruction, we also use the second order derivative filters $f_3 = [-1,2,-1]$, $f_4 = [-1,2,-1]^{\mathrm{T}}$ and $f_3 = [1,-1;-1,1]$. Comparison with the same settings was carried out with the method in [2].

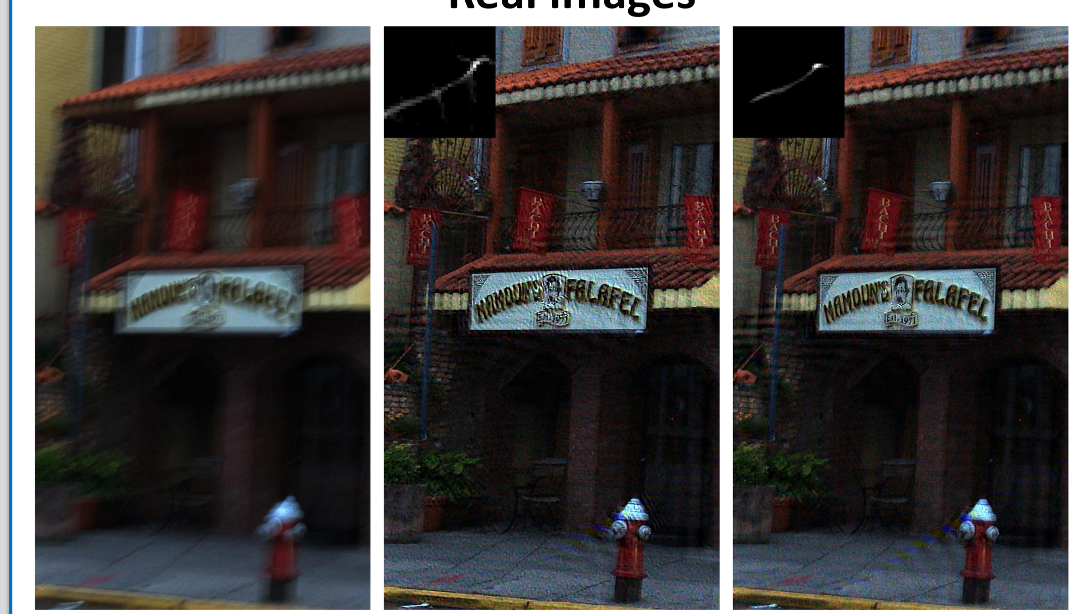
PSNR values for the Y band

image	method	kernel 1	2	3	4	5	6
1	Proposed	**31.04**	30.56	31.18	**33.37**	**31.35**	**30.60**
	Zhou [9]	29.24	**32.92**	**32.27**	30.99	30.64	27.20
2	Proposed	**31.03**	30.99	**31.41**	**32.63**	**30.81**	**32.04**
	Zhou [9]	30.11	**31.00**	30.29	29.25	30.40	31.62
3	Proposed	29.55	**31.07**	30.64	**31.28**	29.20	24.47
	Zhou [9]	**30.77**	29.70	**31.03**	30.18	**29.52**	**30.18**
4	Proposed	**31.04**	**30.47**	**31.75**	**31.66**	**30.31**	**31.21**
	Zhou [9]	30.08	29.96	30.60	30.15	29.48	30.44



Observed | Zhou [2] | Proposed

Real images



Observed | Zhou [2] | Proposed