Variational Bayes for Blind Image Deconvolution

Antonie Brožová

20. 9. 2020

Convolution

$$\boldsymbol{D} = \boldsymbol{k} \circledast \boldsymbol{X} + \boldsymbol{n},$$

- $\boldsymbol{D}$ denotes blurred image
- $\boldsymbol{k}$ denotes convolutional kernel (PSF)
- $\boldsymbol{X}$ denotes sharp image
- $\boldsymbol{n}$ denotes noise

Bayes theorem

$$p(\boldsymbol{\theta}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{\theta})p(\boldsymbol{\theta}),$$

- $p(\boldsymbol{\theta}|\boldsymbol{d})$ denotes posterior distribution
- $p(\boldsymbol{d}|\boldsymbol{\theta})$ denotes distribution od noise
- $p(\boldsymbol{\theta})$ denotes prior distribution
- $\boldsymbol{d}$ denotes data
- $\boldsymbol{\theta}$ denotes latent variables and parameters

Variational Bayes

Approximation of posterior

$$p(\boldsymbol{\theta}|\boldsymbol{d}) \sim q(\boldsymbol{\theta}|\boldsymbol{d}),$$
$$q(\boldsymbol{\theta}|\boldsymbol{d}) = \prod_i q(\theta_i|\boldsymbol{d})$$

Kullback-Leibler divergence of q from p

$$KL\left(q(\boldsymbol{\theta}|\boldsymbol{d}) \parallel p(\boldsymbol{\theta}|\boldsymbol{d})\right) = \mathbb{E}_{q(\boldsymbol{\theta}|\boldsymbol{d})}\left[\ln \frac{q(\boldsymbol{\theta}|\boldsymbol{d})}{p(\boldsymbol{\theta}|\boldsymbol{d})}\right]$$

Kernel and Image Estimates

- Distributions were chosen to be conjugate
- Posterior of image and blur is gaussian $\rightarrow$ their estimates are mean values of the gaussian distributions

Iterative Variational Bayes Algorithm

- Conjugate distributions lead to a set of linear equations for parameters of posteriors
- Iteratively minimizes KL divergence
- Fast, but requires linear system of equations

ELBO - Evidence Lower Bound

$$KL\left(q(\boldsymbol{\theta}|\boldsymbol{d}) \parallel p(\boldsymbol{\theta}|\boldsymbol{d})\right) = \mathbb{E}_{q(\boldsymbol{\theta}|\boldsymbol{d})}\left[\ln \frac{q(\boldsymbol{\theta}|\boldsymbol{d})}{p(\boldsymbol{\theta},\boldsymbol{d})}\right] + p(\boldsymbol{d}) =$$

$$= -\mathcal{L} + p(\boldsymbol{d})$$

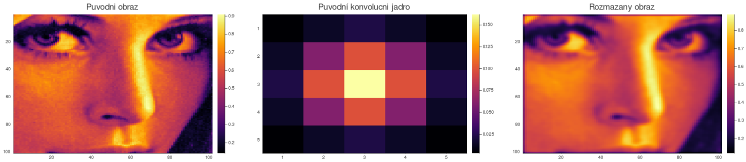
$$\mathcal{L} = p(\boldsymbol{d}) - KL\left(q(\boldsymbol{\theta}|\boldsymbol{d}) \parallel p(\boldsymbol{\theta}|\boldsymbol{d})\right)$$

- Max ELBO = Min KL
- Stochastic gradient descent - slow, but does not require the linear system

Reparametrization trick

- $\mathcal{L}$ requires knowledge of expectation w.r.t. $q(\boldsymbol{\theta}|\boldsymbol{d})$
- Reparametrization $\boldsymbol{\theta} = g_\theta(\epsilon, \dots), \epsilon \sim p(\epsilon)$
- $\nabla_\theta \mathbb{E}_{q(\boldsymbol{\theta}|\boldsymbol{d})}\left[f(\boldsymbol{\theta}, \boldsymbol{d})\right] \approx \frac{1}{L}\sum_{l=1}^{L} \nabla_\theta \left(f(g_\theta(\epsilon^l, \dots), \boldsymbol{d})\right)$

Sharp image and blur

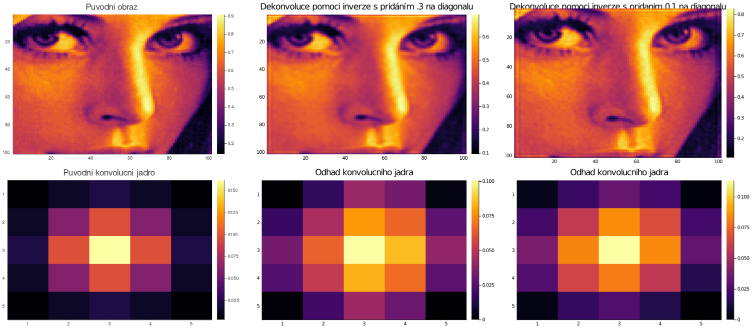


- Posterior $p(\boldsymbol{k}|\boldsymbol{d}) = \mathcal{N}\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right)$
- Image is assumed to be smooth

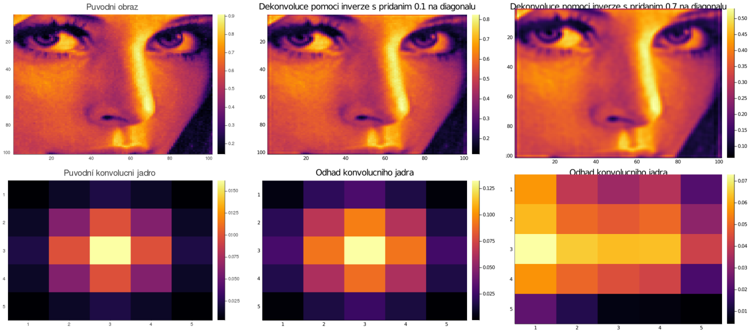
Comparison of two algorithms

1. fully IVB
2. IVB for all variables except for $\boldsymbol{\mu}_k$ a $\boldsymbol{\Sigma}_k$, that are estimated by SGD with reparametrization trick

VB and VB+ELBO for diagonal covariance



VB and VB+ELBO for full covariance



PSNR

Covariance	Blur	SNR	PSNR VB	PSNR ELBO
full	gauss	20	25.93823	25.12414
full	gauss	30	29.91891	28.2933
full	gauss	40	32.58748	29.25239
full	gauss	50	36.51204	29.01416
diag	gauss	20	25.67324	25.91173
diag	gauss	30	30.25197	29.96938
diag	gauss	40	33.04796	32.57338
diag	gauss	50	33.76192	34.88439

Thank you for your attention!