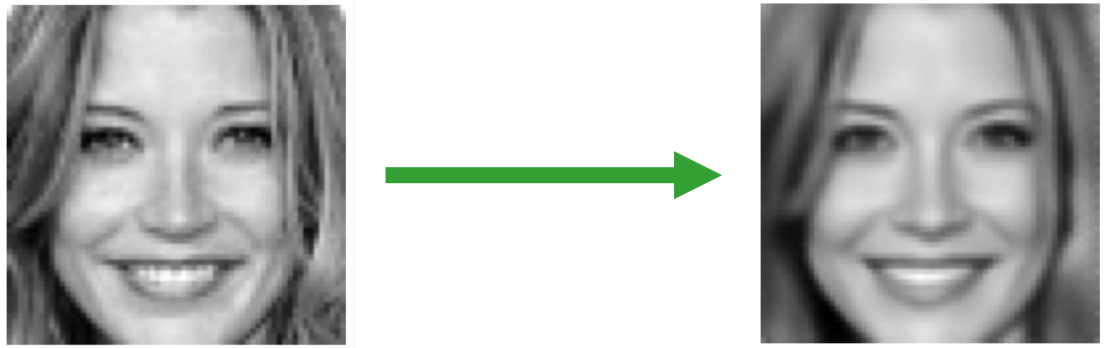
Structured Uncertainty Prediction Networks

[Dorta et al. CVPR 2018]

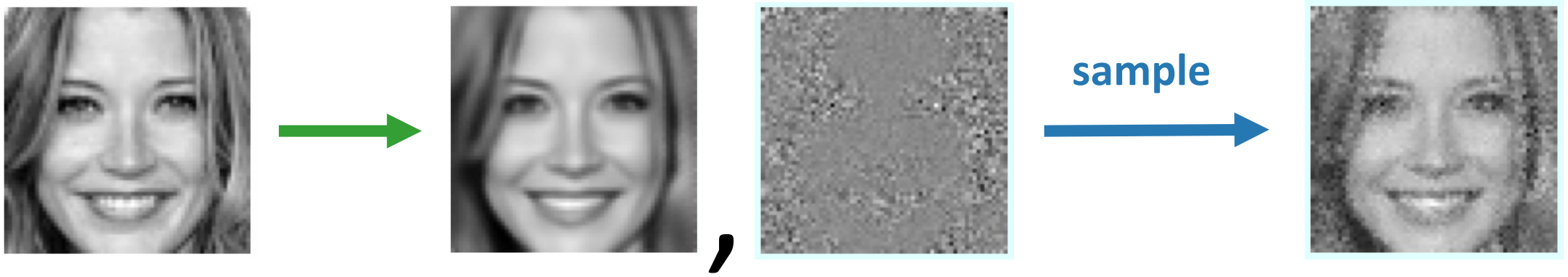
Problem: VAEs produce overly smooth output..



- Fails to capture all the **details** in the data
- Factorised Gaussian (e.g. L2 or diagonal loss) deals with the failures by averaging them across pixels (smoothing)

Problems with the diagonal noise model..

- Factorised noise assumption does not hold for **images**
- Seen by sampling from the likelihood (e.g. diagonal Gaussian)…



reconstruction (mean), diagonal variance

draw from noise model

- The random draw does not match the data

What if we use a structured noise model?

- Instead predict a structured covariance matrix (from latent space)
- We can draw samples to compare…

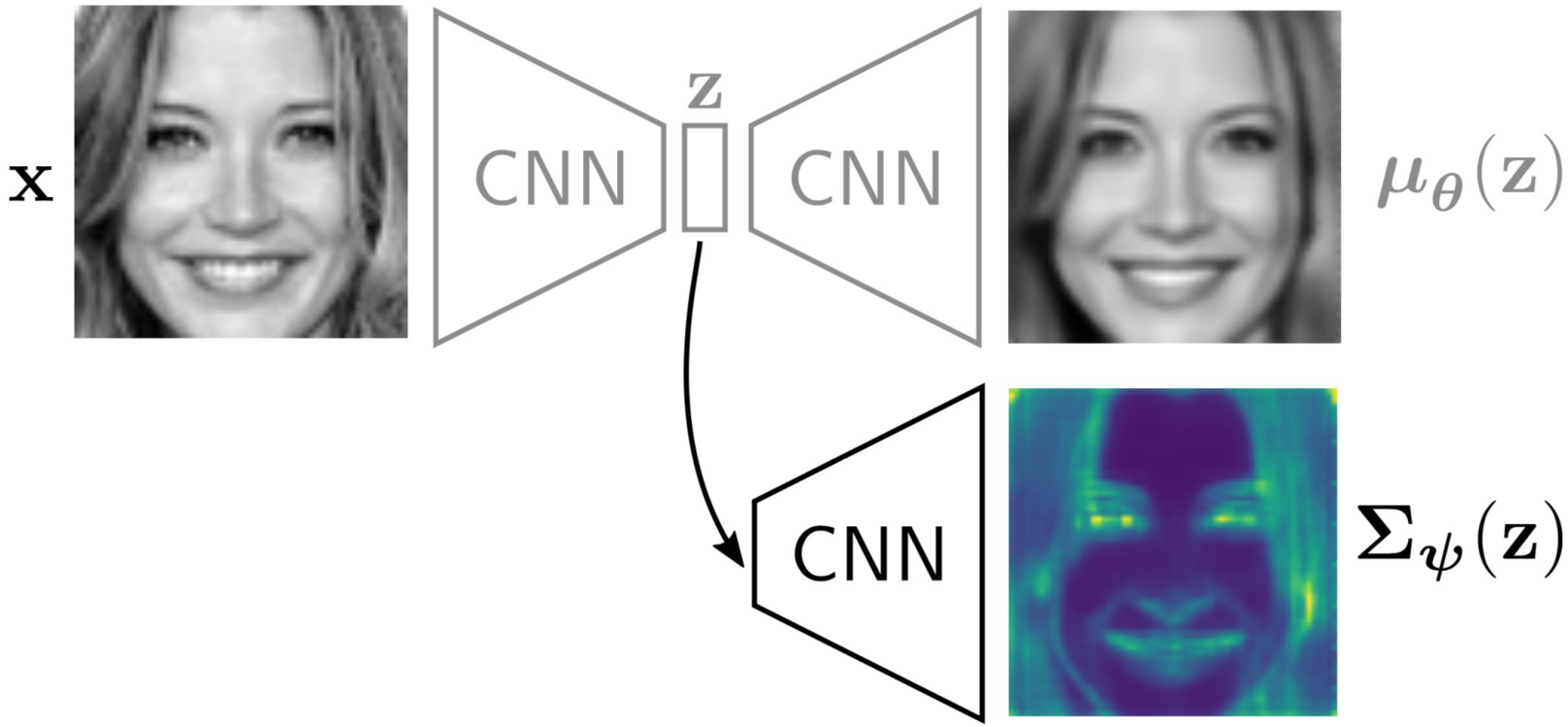
reconstruction (mean), structured covariance

draw from structured noise model

- A random draw captures the statistics of the input data!

Structured uncertainty prediction network

$$p_{\boldsymbol{\theta}}(\mathbf{x} \mid \mathbf{z}) = \mathcal{N}\big(\boldsymbol{\mu}(\mathbf{z}), \boldsymbol{\Sigma}_{\boldsymbol{\psi}}(\mathbf{z})\big).$$

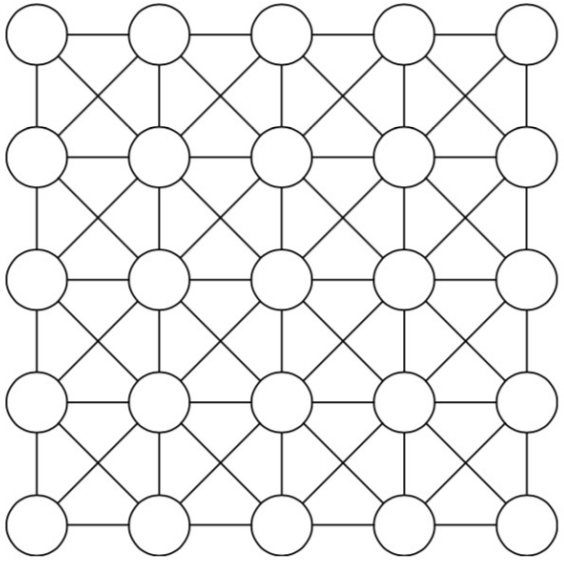


Tractable via sparse connectivity

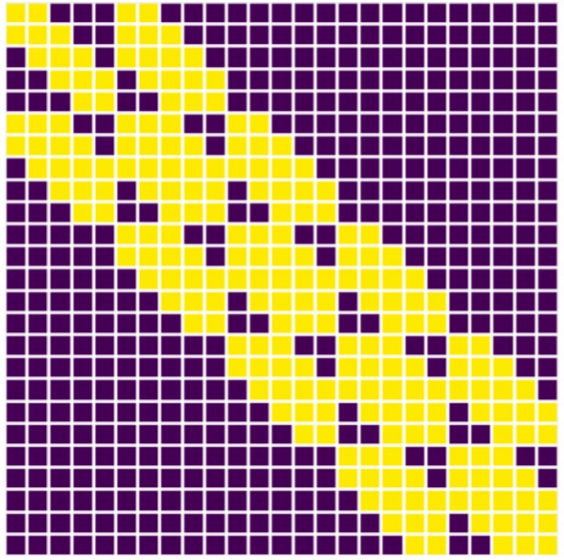
- Parameterise the precision matrix $\boldsymbol{\Sigma}^{-1} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$ for efficiency

$$\min_{\boldsymbol{\psi}} \; \log\left(\left|\boldsymbol{\Sigma}_{\boldsymbol{\psi}}\right|\right) + \left(\mathbf{x}-\boldsymbol{\mu}\right)^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\psi}}\right)^{-1}\left(\mathbf{x}-\boldsymbol{\mu}\right)$$

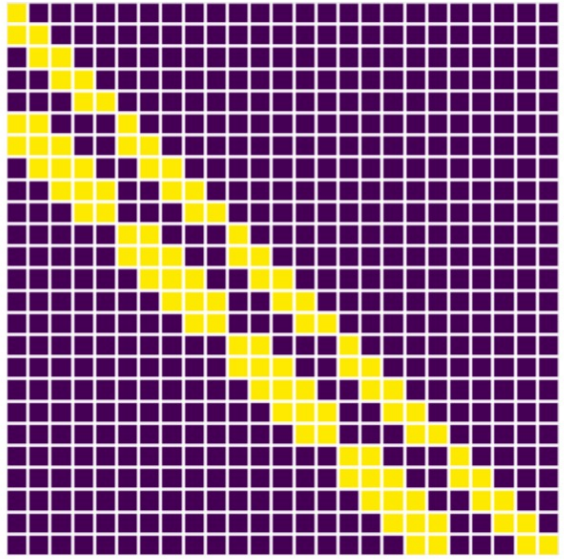
Connectivity



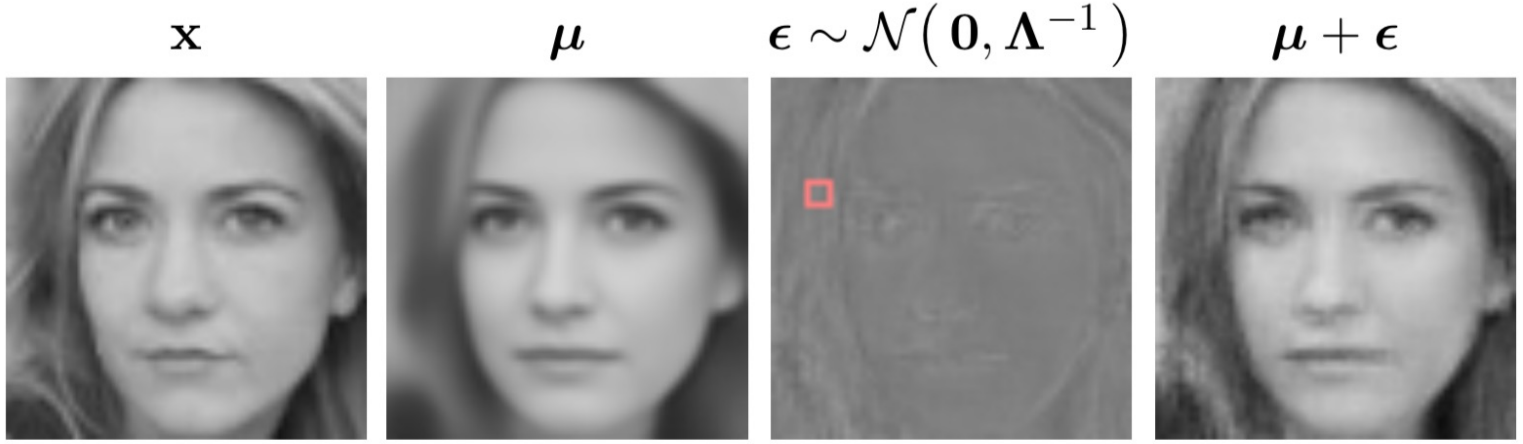
$\boldsymbol{\Sigma}^{-1}$

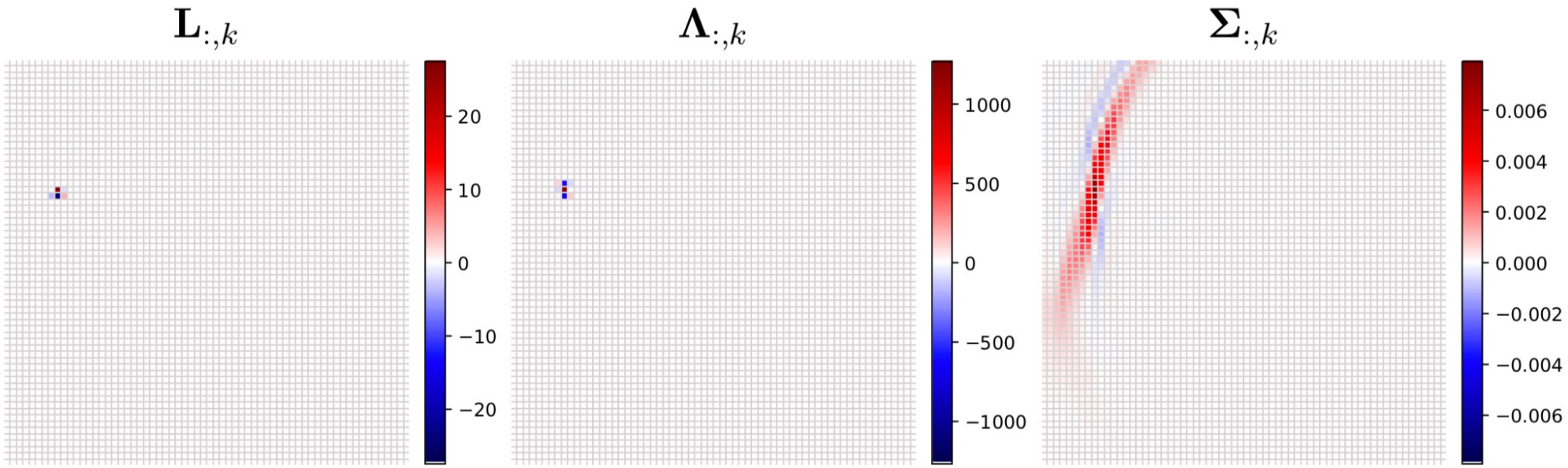


$\mathbf{L}$

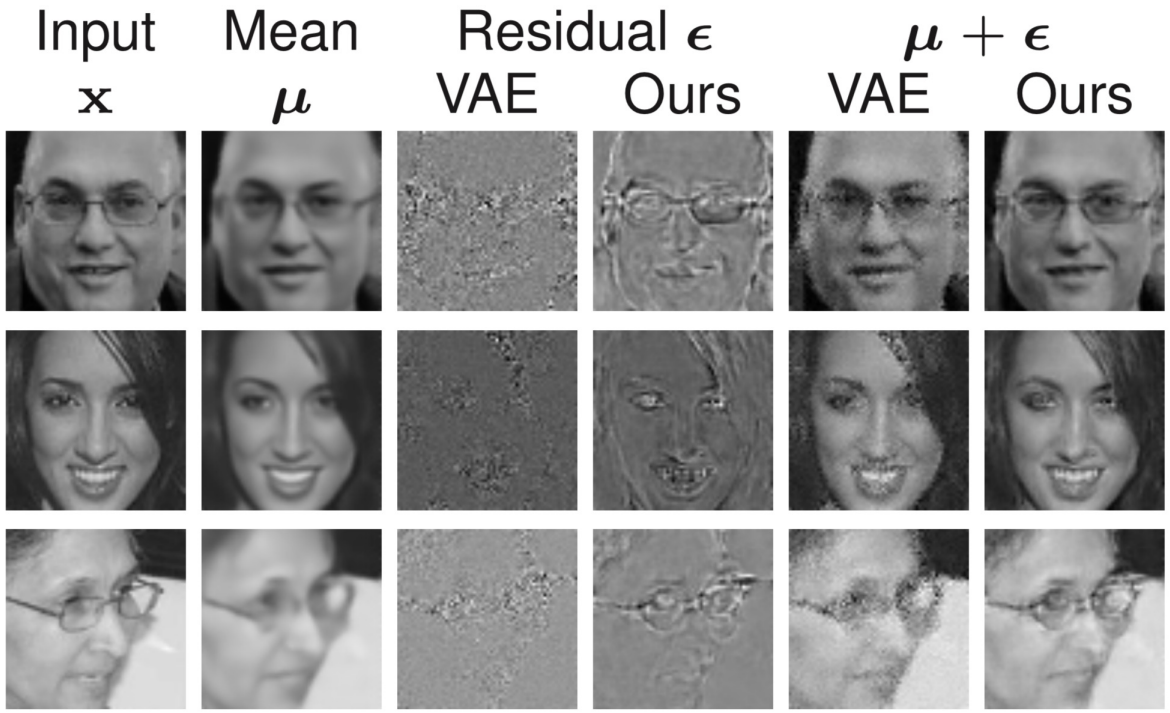


Long range correlations from sparse precision..



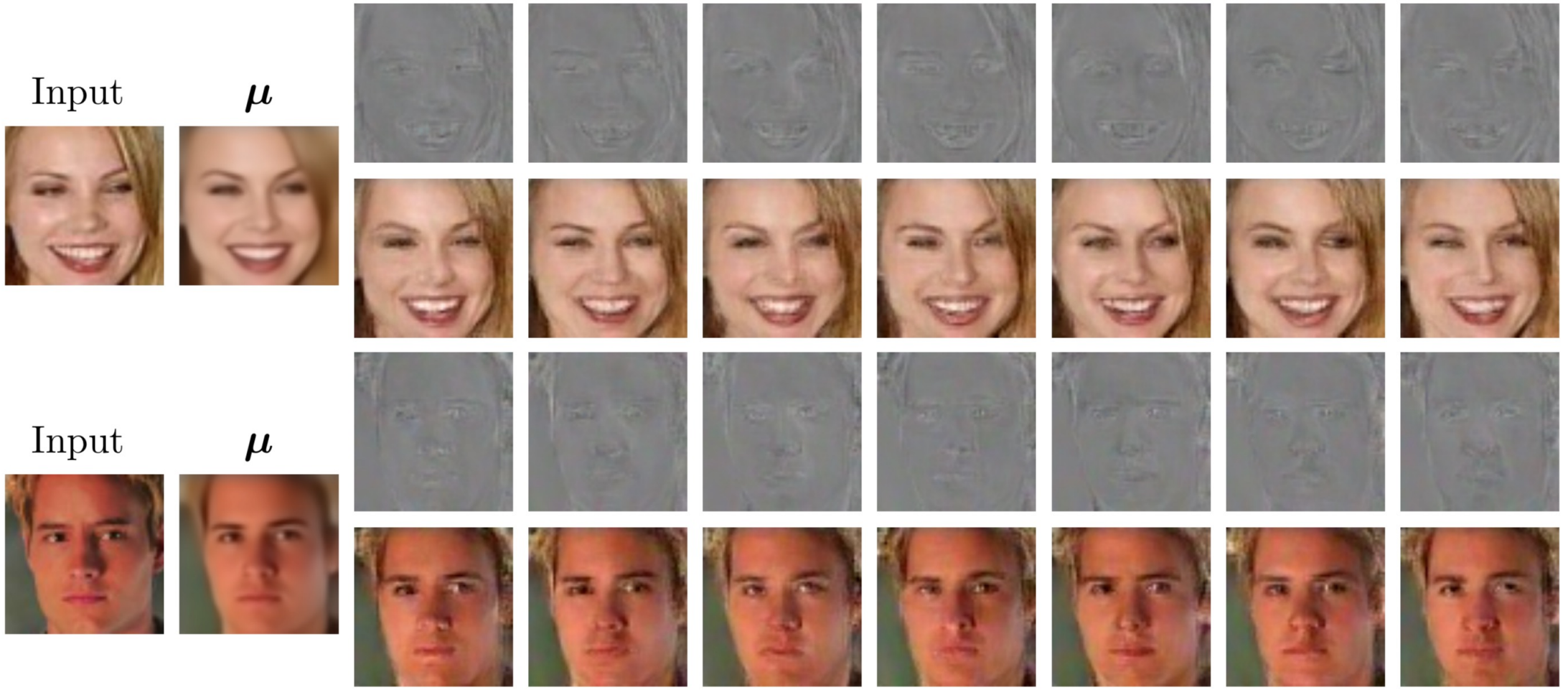


Reconstructions on celebA dataset..

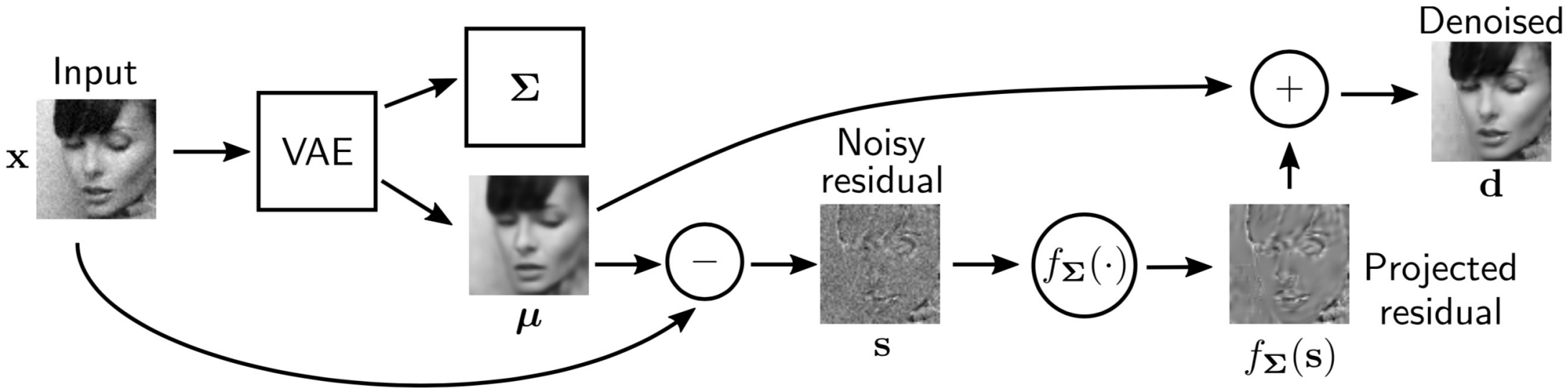


Model	NLL	$-\log p(\mathbf{x} \mid \mathbf{z})$
VAE [1]	-5378 ± 931	-6079 ± 936
Ours	-7753 ± 1323	-8386 ± 1339

Reconstruction variation

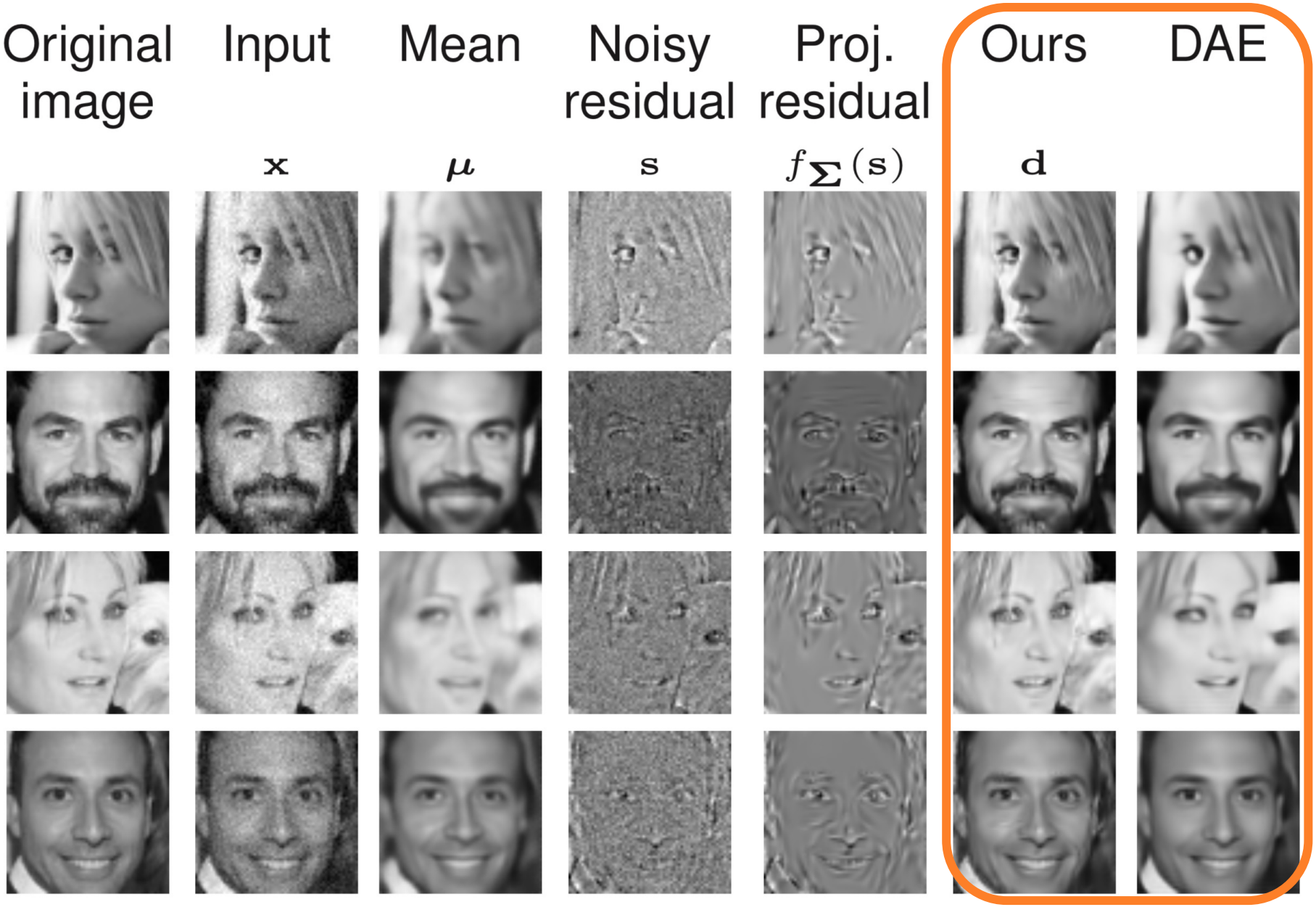


What about the noisy projection evaluation?



Model	**MSE**	**PSNR**	**SSIM**
DAE	0.005 ± 0.003	28.89 ± 1.69	0.90 ± 0.03
Ours	$\mathbf{0.003 \pm 0.001}$	$\mathbf{31.38 \pm 0.92}$	$\mathbf{0.92 \pm 0.02}$

But what about the noisy projection evaluation..



Limitations

- Lack of proper predictive posterior in the VAE latent space
- Difficult to know where to draw samples from:

Limitations

- Standard NN caveats apply..
  - What happens away from the training data?
  - Constraints on the function?
  - True epistemic uncertainty?
- Likelihood function only works for the category trained on..