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

- 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...

• A random draw captures the statistics of the input data!
Structured uncertainty prediction network

\[ p_\theta(x \mid z) = \mathcal{N}(\mu(z), \Sigma_\psi(z)) \]
Tractable via sparse connectivity

- Parameterise the precision matrix $\Sigma^{-1} = LL^T$ for efficiency

$$\min_{\psi} \log \left( |\Sigma_\psi| \right) + (x - \mu)^T (\Sigma_\psi)^{-1} (x - \mu)$$
Long range correlations from sparse precision..
Reconstructions on celebA dataset.

| Model   | NLL         | $- \log p(x \mid z)$ |
|---------|-------------|-----------------------|
| VAE [1] | $-5378 \pm 931$ | $-6079 \pm 936$       |
| Ours    | $-7753 \pm 1323$ | $-8386 \pm 1339$      |
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   | 0.003 ± 0.001 | 31.38 ± 0.92 | 0.92 ± 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..