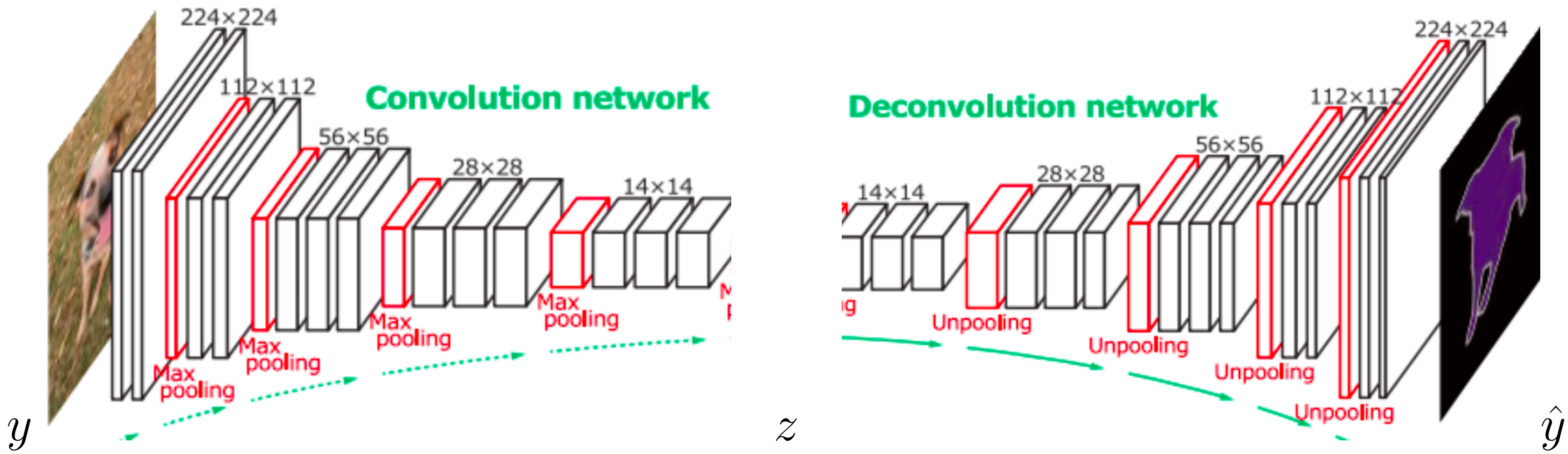
TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Gaussian Noisy Channel RDAs

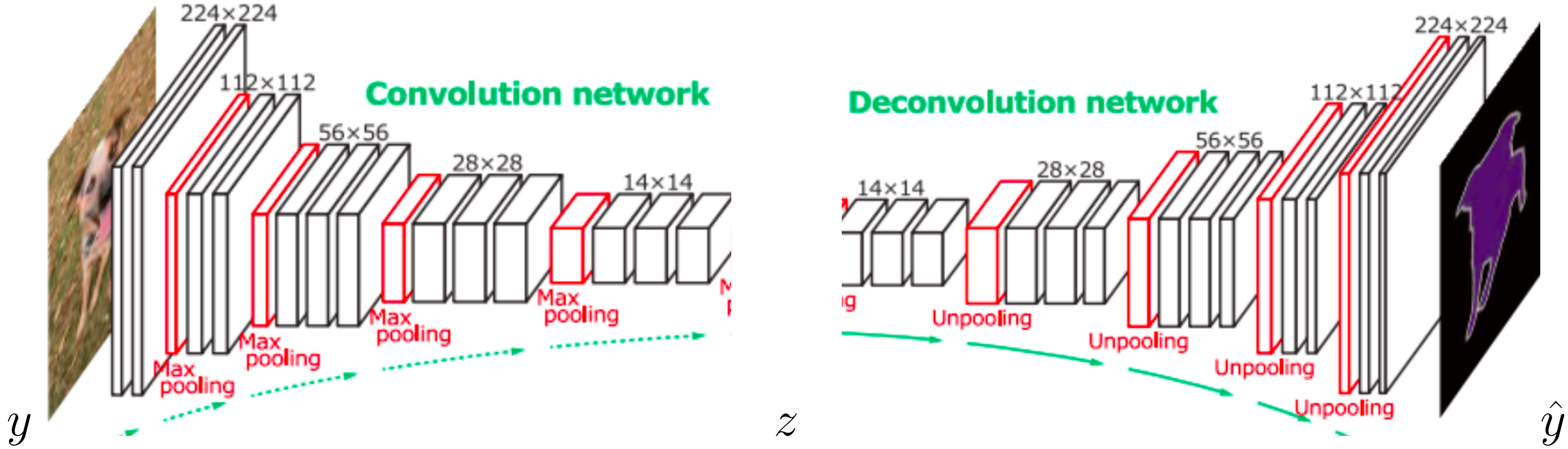
The Noisy Channel RDA

$$\Phi^* = \underset{\Phi}{\operatorname{argmin}} E_{y,\epsilon} \ln \frac{p_\Phi(z_\Phi(y,\epsilon)|y)}{\hat{p}_\Phi(z_\Phi(y,\epsilon))} + \lambda \mathrm{Dist}(y, y_\Phi(z_\Phi(y,\epsilon)))$$



We can require $\hat{p}_\Phi(z)$ be Gaussian. In that case we can sample z from $\hat{p}_\Phi(z)$ and generate images (as in a GAN).

A General Autoencoder



We show below that for $p_\Phi(z|y)$ and $\hat{p}_\Phi(z)$ both required to be Gaussian we can assume without loss of generality that

$$\hat{p}_\Phi(z) = \mathcal{N}(0, I)$$

Sampling

Sample $z \sim \mathcal{N}(0, I)$ and compute $y_\Phi(z)$

[Alec Radford]

This is **sampling** — not compression. This is "decompressing" noise.

Gaussian Noisy-Channel RDA

We now show that a reparameterization can always convert $\hat{p}_\Phi(z)$ to a zero-mean identity-covariance Gaussian.

$$\Phi^* = \operatorname*{argmin}_\Phi E_{y,\epsilon}\ \ln \frac{p_\Phi(z_\Phi(y,\epsilon)|y)}{\hat{p}_\Phi(z_\Phi(y,\epsilon))} + \lambda \mathrm{Dist}(y, y_\Phi(z_\Phi(y,\epsilon)))$$

$$\begin{aligned}
z_\Phi(y,\epsilon) &= \mu_\Phi(y) + \sigma_\Phi(y) \odot \epsilon \quad \epsilon \sim \mathcal{N}(0, I) \\
p_\Phi(z[i]|y) &= \mathcal{N}(\mu_\Phi(y)[i], \sigma_\Phi(y)[i])) \\
\hat{p}_\Phi(z[i]) &= \mathcal{N}(\hat{\mu}_z[i], \hat{\sigma}_z[i]) \\
\mathrm{Dist}(y, \hat{y}) &= ||y - \hat{y}||^2
\end{aligned}$$

Gaussian Noisy-Channel RDA

$$\Phi^* = \underset{\Phi}{\operatorname{argmin}}\, E_{y,\epsilon}\ \ln \frac{p_\Phi(z_\Phi(y,\epsilon)|y)}{\hat{p}_\Phi(z_\Phi(y,\epsilon))} + \lambda \mathrm{Dist}(y, y_\Phi(z_\Phi(y,\epsilon)))$$

We will show that we can fix $\hat{p}_\Phi(z)$ to $\mathcal{N}(0, I)$.

$$\begin{aligned} p_\Phi(z[i]|y) &= \mathcal{N}(\mu_\Phi(y)[i], \sigma_\Phi(y)[i]) \\ \hat{p}_\Phi(z[i]) &= \mathcal{N}(0, 1) \\ \mathrm{Dist}(y, \hat{y}) &= ||y - \hat{y}||^2 \end{aligned}$$

Gaussian Noisy-Channel RDA

$$\Phi^* \;=\; \operatorname*{argmin}_{\Phi} E_{y,\epsilon}\;\ln\frac{p_\Phi(z_\Phi(y,\epsilon)|y)}{\hat{p}_\Phi(z_\Phi(y,\epsilon))} + \lambda \mathrm{Dist}(y, y_\Phi(z_\Phi(y,\epsilon)))$$

$$=\; \operatorname*{argmin}_{\Phi}\; E_{y\sim\mathrm{Pop}}\left(\begin{array}{c} KL(p_\Phi(z|y), \hat{p}_\Phi(z)) \\ \\ +\lambda\; E_\epsilon\; \mathrm{Dist}(y,\; y_\Phi(z_\Phi(y,\epsilon))) \end{array}\right)$$

Closed Form KL-Divergence

$$KL(p_\Phi(z|y), \hat{p}_\Phi(z))$$

$$= \sum_i \frac{\sigma_\Phi(y)[i]^2 + (\mu_\Phi(y)[i] - \mu_z[i])^2}{2\sigma_z[i]^2} + \ln \frac{\sigma_z[i]}{\sigma_\Phi(y)[i]} - \frac{1}{2}$$

Standardizing $\hat{p}_\Phi(z)$

$$KL(p_\Phi(z|y), p_\Phi(z))$$

$$= \sum_i \frac{\sigma_\Phi(y)[i]^2 + (\mu_\Phi(y)[i] - \mu_z[i])^2}{2\sigma_z[i]^2} + \ln \frac{\sigma_z[i]}{\sigma_\Phi(y)[i]} - \frac{1}{2}$$

$$KL(p_{\Phi'}(z|y), \mathcal{N}(0, I))$$

$$= \sum_i \frac{\sigma_{\Phi'}(y)[i]^2 + \mu_{\Phi'}(y)[i]^2}{2} + \ln \frac{1}{\sigma_{\Phi'}(y)[i]} - \frac{1}{2}$$

Standardizing $\hat{p}_\Phi(z)$

$$
\begin{aligned}
KL_\Phi &= \sum_i \frac{\sigma_\Phi(y)[i]^2 + (\mu_\Phi(y)[i] - \mu_z[i])^2}{2\sigma_z[i]^2} + \ln \frac{\sigma_z[i]}{\sigma_\Phi(y)[i]} - \frac{1}{2} \\
KL_{\Phi'} &= \sum_i \frac{\sigma_{\Phi'}(y)[i]^2 + \mu_{\Phi'}(y)[i]^2}{2} + \ln \frac{1}{\sigma_{\Phi'}(y)[i]} - \frac{1}{2}
\end{aligned}
$$

Setting Φ' so that

$$
\begin{aligned}
\mu_{\Phi'}(y)[i] &= (\mu_\Phi(y)[i] - \mu_z[i])/\sigma_z[i] \\
\sigma_{\Phi'}(y)[i] &= \sigma_\Phi(y)[i]/\sigma_z[i]
\end{aligned}
$$

gives $KL(p_\Phi(z|y), \hat{p}_\Phi(z)) = KL(p_{\Phi'}(z|y), \mathcal{N}(0, I))$.

END