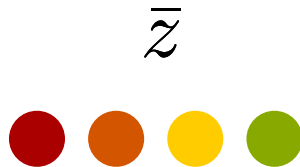




Latent filters  
 $(k, m, n)$

Average  
filters  
→



Latent vector  
of size  $k$

Compute  
loss  
→

$$\frac{\beta|z|}{k} D_{KL}(\mathcal{N}(\bar{\mu}, \bar{\sigma}^2) || p(z))$$

KL loss term