Supplement: Scene Classification with Semantic Fisher Vectors

1. Direct Implementation of a Semantic Fisher Vector

We follow the derivations in Appendix A of [2] to compute the Fisher information matrix for a Dirichlet Mixture distribution. For a K mixture model with mixture weights w_s and component parameters α_s, the following was shown to be a reasonable assumption.

$$\begin{aligned}\frac{\partial p(k|x)}{\partial \alpha_s} &= p(k|x)\left[\delta(s,k) - p(k|x)\right]\frac{\partial \log p(x|s)}{\partial \alpha_s} \\ &\approx 0\end{aligned} \quad (1)$$

where $\delta(s,k)$ is an indicator function, that equals 1 if $s=k$, and is 0 otherwise. This assumption is valid for large mixture distributions, where the posterior probabilities $P(k|x)$ are very peaky, and therefore $P(k|x)P(s|x) \approx 0$ if $k \neq s$ and $P(k|x) \approx P(k|x)P(s|x)$ if $k = s$.

Second order derivative of a Dirichlet mixture log likelihood with respect to its parameters can be expressed as,

$$\frac{\partial^2 \mathcal{L}}{\partial \alpha_{sm}\partial \alpha_{kl}} = \frac{\partial}{\partial \alpha_{sm}} p(k|\pi)\left(\psi(\sum_l \alpha_{kl}) - \psi(\alpha_{kl}) + \log \pi_l\right)$$

$$\begin{aligned}&= \left(\frac{\partial p(k|\pi)}{\partial \alpha_{sm}}\right)\left(\psi(\sum_l \alpha_{kl}) - \psi(\alpha_{kl})\right) \\ &+ p(k|\pi)\left(\frac{\partial}{\partial \alpha_{sm}}\psi(\sum_l \alpha_{kl}) - \frac{\partial}{\partial \alpha_{sm}}\psi(\alpha_{kl})\right) \\ &= 0 + p(k|\pi)\left(\psi'(\sum_l \alpha_{kl}) - \psi'(\alpha_{kl})\delta(l,m)\right)\delta(k,s)\end{aligned} \quad (2)$$

where π_l is the l^{th} dimension of the data point π, which is a probability vector, and $\psi'(x) = \frac{\partial \psi(x)}{\partial x}$ is a digamma function. The presence of $\delta(k,s)$ in the expression indicates that $\frac{\partial^2 \mathcal{L}}{\partial \alpha_{sm}\partial \alpha_{kl}} = 0$ if $k \neq s$, that is, if the gradient is with respect to parameters of two different mixture components. The Fisher Information matrix, therefore simplifies into the following block diagonal form.

$$\begin{aligned}\mathcal{F}_{lm} &= E\left[-\frac{\partial^2 \log P(\pi|\{\alpha_k, w_k\}_{k=1}^K)}{\partial \alpha_{kl}\partial \alpha_{km}}\right] \\ &= E\left[p(k|\pi)\right]\left(\psi'(\alpha_{kl})\delta(l,m) - \psi'(\sum_l \alpha_{kl})\right) \\ &= w_k\left(\psi'(\alpha_{kl})\delta(l,m) - \psi'(\sum_l \alpha_{kl})\right)\end{aligned} \quad (3)$$

The matrix $\mathcal{F}^{-1/2}$ is used to scale the DMM based Fisher scores of each image BoS (see eq (8) in the paper). The resulting image representation is referred to as a Dirichlet mixture Fisher vector (DMM FV). Note that we do not use Fisher gradients of mixture weights in our image representation, as they often result in negligible performance gains [1].

References

[1] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. In *Proceedings of the 11th European conference on Computer vision: Part IV*, ECCV'10, pages 143–156, Berlin, Heidelberg, 2010. Springer-Verlag. 1

[2] J. Sánchez, F. Perronnin, T. Mensink, and J. J. Verbeek. Image classification with the fisher vector: Theory and practice. *International Journal of Computer Vision*, 105(3):222–245, 2013. 1