Sketching with Style: Visual Search with Sketches and Aesthetic Content

J. Collomosse, T. Bui, M. Wilber, C. Fang, H. Jin

October 20, 2018

- Visual similarity based on structural constraints (query sketch image) and aesthetic or style constraints (one or more contextual images).
- Style network is based on triplet (Siamese) network with training image a, hard positive image p, and hard negative image n. Loss to be minimized is

$$\mathcal{L}(a,p,n) = [m + |f(a) - f(p)|^2 - |f(a) - f(n)|^2]_+$$

  Helps decorrelate objects with other meanings, for *eg.* skulls and scary scenes.
- Read *Generalisation and sharing in triplet convnets for sketch based visual search* by T. Bui *et al.* for structure network.
- For training sketch s with anchor a on positive p and negative n: Run s on style, a on structure, p and n on both, and obtain concatenation $256D$ vector. Similar loss with regularization as before:

$$\mathcal{L}'(a,p,n) = \Sigma_{i\in\{a,p,n\}}\phi_s\mathcal{S}(i) + \phi_t\mathcal{L}(a,p,n)$$

- For testing of sketch s and anchor images a_i, find

$$d = h(\textbf{concatenation}(\text{style}(s),\ \Sigma_i\ w_i\ \text{structure}(a_i)))$$