Large-scale long tailed recognition in open-world.

Dynamic meta-embedding:

v^{direct} → feature from the input image. (layer before final layer)

Depends on

v^{memory} → relate visual concept in memory module.

v^{meta} → sent to the final classification layer

Learning visual memory: $\{c_i\}$

class centroid in minibatch.

$$v^{memory} = o^T M := \sum_{i=1}^{m} o_i c_i$$

$o \in \mathbb{R}^k$ hallucination coefficient.

$$o = T_{hal}(v^{direct})$$

Element wise operation (multi)

meta ... direct ... memory]

$v^{meta} = \frac{1}{\gamma}\left(v^{direct} + e \otimes v \right)$

reachability

$\hat{y} := \min_i \| v^{direct} - c_i \|_2$ closest class.

concept Selector:

$e = \tanh\left(T_{sel}\left(v^{direct} \right) \right)$

modulated Attention:

$$f^{att} = f + mA(f) \otimes SA(f)$$

final loss function

$$L = \sum_{n=1}^{N} L_{CE}\left(v_n^{meta}, y_n \right) + \lambda L_m\left(v_n^{meta}, \{c_i\}_{i=1}^{K} \right)$$