Non-local Neurons

Generic Non-local operation $y_i = \frac{1}{C(x)} \sum_{\forall j} f(x_i, x_j) g(x_j)$

weighted sum

step 1

Normalization

output position

Instantiation: Design choice of the f & g

$g(x_j) = W_g x_j$ // [1×1 convolution or 1×1×1]

Gaussian: $f(x_i, x_j) = e^{x_i^T x_j}$

$C(x) = \sum_{\forall j} f(x_i, x_j)$

embedded gaussian: $f(x_i, x_j) = e^{\theta(x_i)^T \phi(x_j)}$ → $W_\theta x_i$, → $W_\phi(x_j)$

* Generalization of Non local

Dot Product: $f(x_i, x_j) = \theta(x_i)^T \phi(x_j)$

Concentration: $f(x_i, x_j) = \text{ReLU}(W_f^T [\theta(x_i), \phi(x_j)])$

* Wrapping Everything in Non local box

$z_i = W_z y_i + x_i$ // step 2