**ABSTRACT**

An increasing number of computer vision tasks can be tackled with deep features, which are the intermediate outputs of a pre-trained Convolutional Neural Network. Despite the astonishing performance, deep features extracted from low-level neurons are still below satisfaction, arguably because they cannot access the spatial content contained in the higher layers. In this paper, we present InterActive, a novel algorithm which computes the activeness of the neurons at each level, and extracts information from the last layer. Activeness is propagated through a neural network in a top-down manner, carrying high-level context and improving the descriptive power of low-level mid-level neurons. Visualization indicates that neuron activeness can be interpreted as spatial-weighted neuron responses. We achieve state-of-the-art classification performance on a wide range of image datasets.

**THE PROPOSED ALGORITHM**

### Improved deep feature extraction

**Input**: an image with $W \times H$ pixels.

**Reference model**: a pre-trained VGGNet.

Original deep feature extraction: resizing the image into $224 \times 224$, passing it through the network, and extracting the intermediate response on the $t$-th layer.

**Improved deep feature extraction**: resizing the input to $64 \times 64$:

- The aspect ratio is maximally preserved;
- The area (number of pixels) is $512^2$;
- The width and height are multipliers of $32$ (the down-sampling rate of VGGNet) passing the resized image through the network, extracting the intermediate response on the $t$-th layer, and averaging over all the spatial positions.

**Significant accuracy gain**

**Main reason**: conventional deep feature extraction only involves forward-propagation, useful information in high-level layers is discarded.

**Solution**: introducing a back-propagation process into deep feature extraction!

### Inter-layer activeness propagation

**Key idea**: unsupervised back-propagation in the deep feature extraction process.

- Defining a score function at the top level;
- Back-propagating gradients to obtain the activeness of neuron connections;
- Collecting the activeness of neuron connections as the activeness of neurons.

A pre-trained CNN: $h(x^{(0)'}, \theta)$, $x^{(0)}$ is the input, and $\theta$ is the weights. $x^{(0)}$ is the neuron responses on the $t$-th layer, a $W_t \times H_t \times D_t$ cube. $x^{(0)}$ ($D_t$-dimensional) is the average over all spatial positions of $x^{(0)}$.

**Feature distribution on the top layer**

We study the PDF of neuron responses on the $7$-th layer, i.e., the starting point of backprop.

We assume the following PDF:

$$f(x^{(0)}) = C_p \times \exp\left[-\frac{1}{2}(x^{(0)} - \mu)^2\right],$$

where $p$ is the norm (1 or 2) and $C_p$ is the normalization coefficient.

**Score function and activeness of connections**

The activeness of neuron connection is computed using the score function, i.e., the gradient of log-distribution over $\theta^{(0)}$, $\alpha^{(0)} = \frac{\partial \log f(x^{(0)})}{\partial \theta^{(0)}}$, $\alpha^{(0)}$ is the layer score, and $\frac{\partial \log f(x^{(0)})}{\partial \theta^{(0)}}$ is the inter-layer activeness.

**Main problem**:

We use InterActive to update the fc-6 neurons.

**Results**

- **ImageNet**
  - Top 1: 26.3% vs 25.4% (8%)
  - Top 5: 57.5% vs 55.4% (2%)

**REFERENCES**


**ACKNOWLEDGEMENTS**

This work was partly done when Lingxi Xie and Liang Zheng were interns at MSR. They contributed equally. This work was supported by NSF Grant 1536102 and ARO grant W111NF15-1-0290, Faculty Research Gift Awards by NEC Labs of America and Blippar, and NSF CS4242001. We thank John Flynn, Xiao Dong, Jianyu Wang, Jianhua Ma and Zhoutao Zhu for instructive discussions.