INTRODUCTION

- Deep convolutional neural networks (CNNs) appear to be the most plausible computational models of visual object recognition in the brain. CNNs have achieved nearly human-level performance in various computer vision tasks. Moreover, recent studies indicate that internal representations of CNNs are more similar to neural responses than other models of the visual cortex.

- Electroencephalography (EEG) enables us to record local field potentials (LFPs) with high spatiotemporal resolution. LFPs in various frequency bands may contribute to neural representations at mesoscale, complementary to neuronal firing [2]. In the primate visual cortex, specific frequency bands subserve feedforward or feedback processing. However, it has been unclear what kind of visual information such frequency-specific activities represent.

- Do predictions of ECoG responses from CNN features have specificity in the frequency domain?
- How are frequency-specific prediction modulated along CNN layers and time?
- What visual properties do the encoding models explain?

MATERIALS AND METHODS

- Image set
  - Total 12000 natural images (building, body part, face, foliage, fruit, fur, glass, insect, leather, metal, paper, tool)

- Recording neural responses in the primate inferior temporal cortex
  - We recorded cortical potentials of 128 channel electrocorticography (ECoG) covering from macaque posterior ITC to anterior ITC.
  - We computed the amplitude of each frequency (1-500 Hz) by complex Morlet wavelet convolution.
  - We downsampled the amplitude for each time window (20 ms), and then conducted trial averaging.

- Diverse image features from deep convolutional neural networks
  - Deep convolutional neural networks (CNNs) have achieved nearly human-level performance in various computer vision tasks.
  - Higher layers in CNNs have higher-level, more abstract and spatially invariant representations [3].
  - We used a pretrained model of VGGNet-16 [4], which has 13 convolution layers.
  - We extracted outputs at each convolution layer using the same image set.

- Encoding frequency-band specific responses from image features
  - Encoding ECoG features from CNN features by ridge regression (regularized linear regression)
  - An encoding model is specified by one ECoG electrode, time window, frequency, and CNN layer.
  - We first optimized each model with training set, and then evaluated each model’s prediction accuracy with test set.
  - Each model’s prediction accuracy was evaluated as Pearson correlation between predicted and true responses.

- Correspondence between the representations of convolutional neural networks and the activities in inferior temporal cortex measured by electrocorticography

RESULTS

- Specificity of prediction accuracy in the frequency domain
  - An example: prediction from conv1 layer to one electrode and time window

- Layer dependence and temporal modulation of prediction accuracy
  - Predictions for all test images

- Visual representations of each frequency-band estimated by encoding models

SUMMARY

- Neural responses in the primate ITC measured by ECoG were predicted by CNN features in a frequency-specific manner.
- Lower-frequency (theta) activities were better predicted by CNN features from middle or higher layers, whereas higher-frequency (low gamma) activities were predicted equally well from almost all the layers.
- Lower-frequency activities were most well predicted at 300-400ms after stimulus onset, whereas higher-frequency activities were at 50-150ms after stimulus onset.
- Visual representations estimated by the best encoding model of each frequency band indicated frequency-specific representations of visual attributes.

References


Conflict of Interest: we declare no competing financial interest and no conflict of interest.