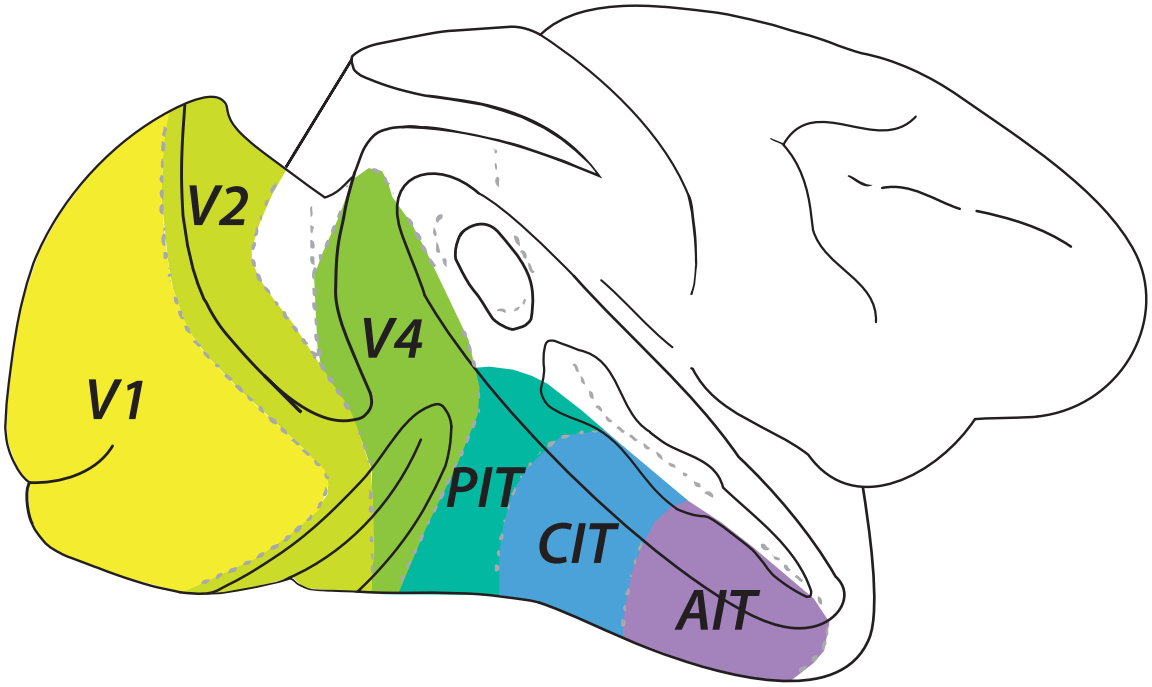
How do CNN units achieve translation invariance?

Dean Pospisil

6/29/17

Region of Interest: **V4**



Network of Interest: AlexNet

fc 8 identification

fc 7

fc 6 fully connected

Max Pooling

conv 5

conv 4

conv 3

Normalization

Max Pooling

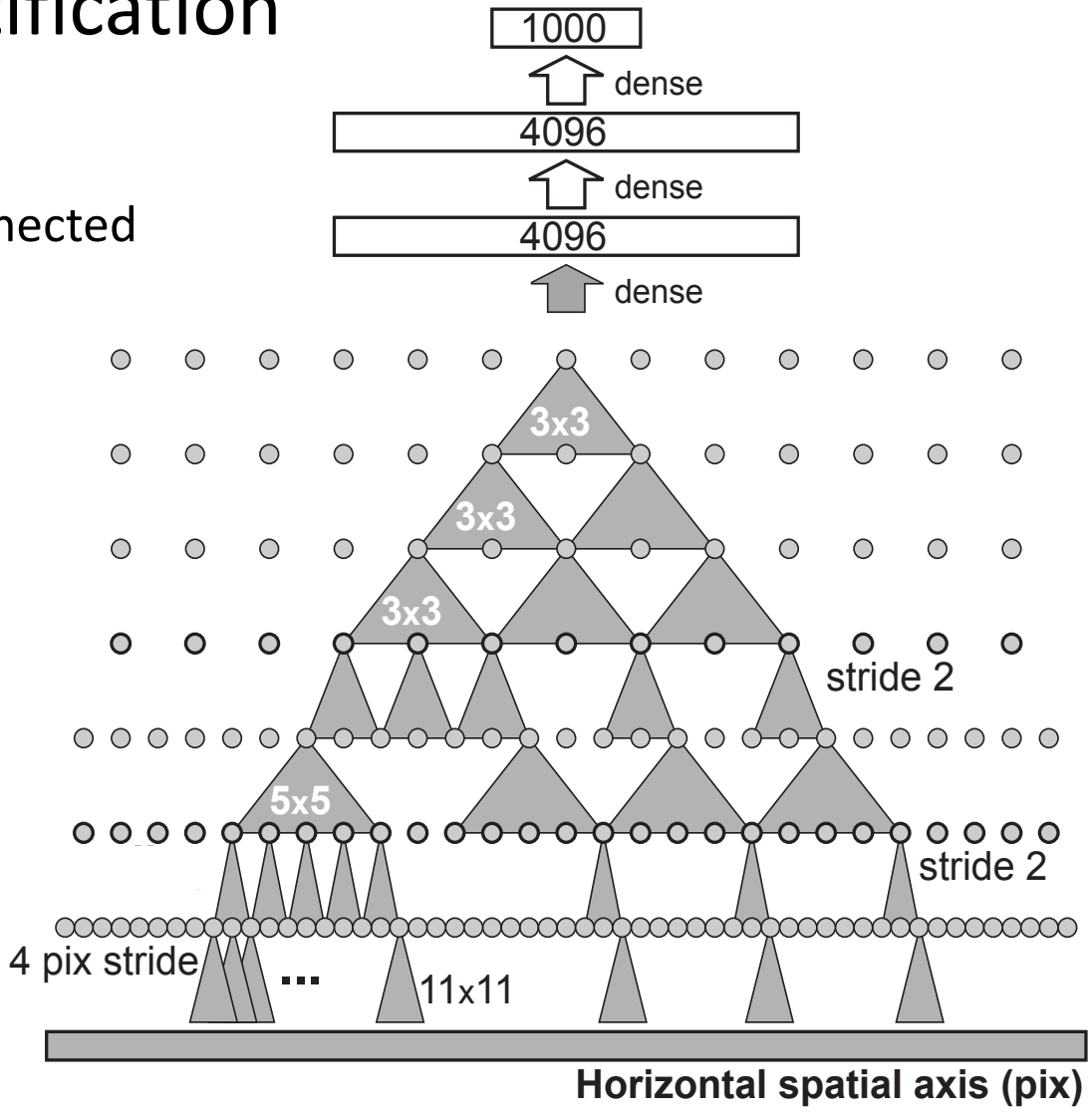
conv 2

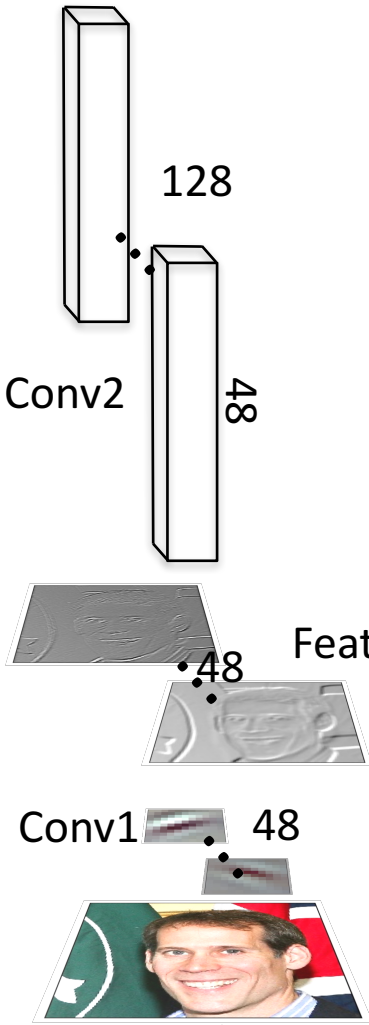
Normalization

Max Pooling

conv 1

convolutional

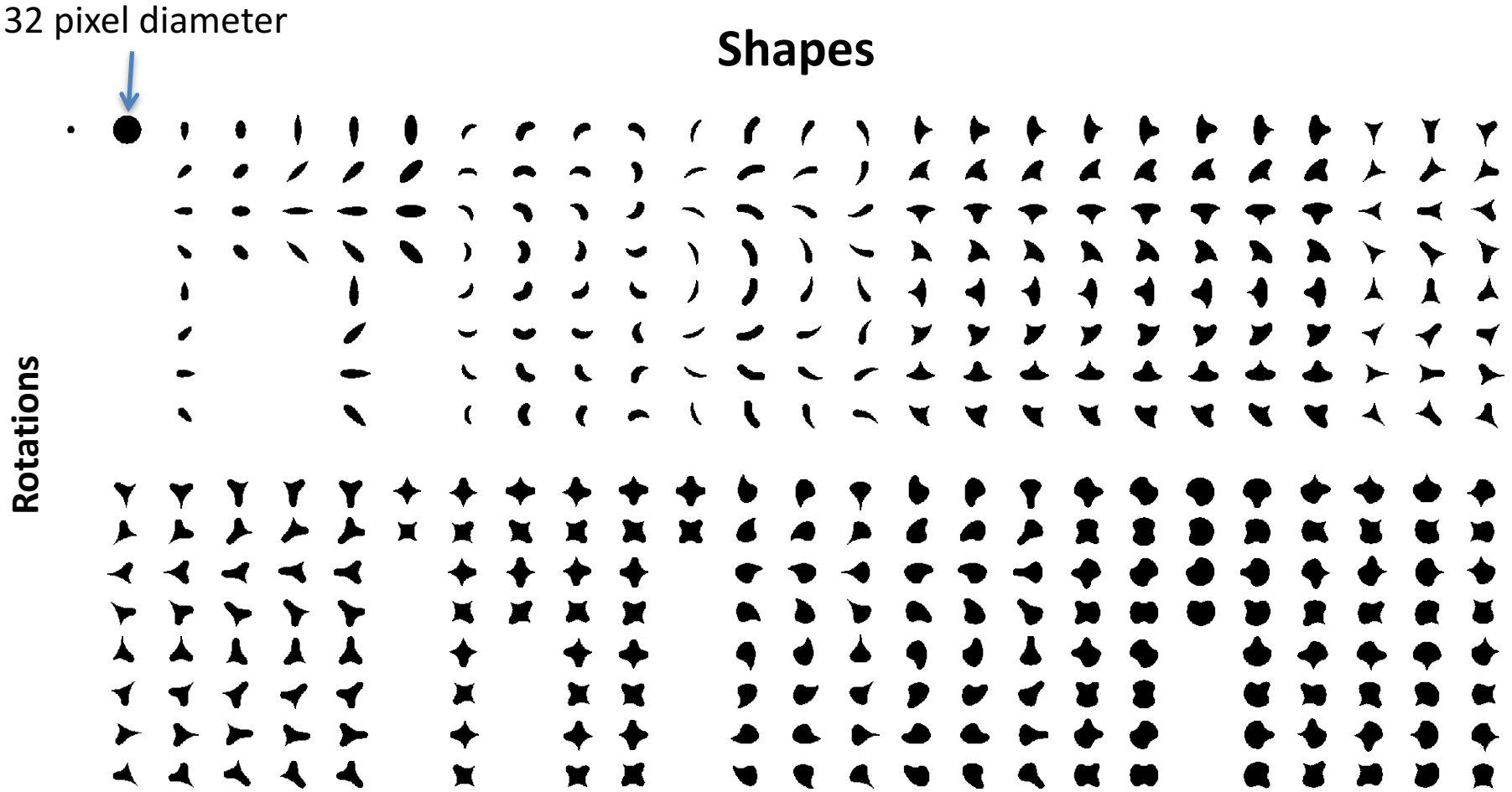




Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton.
"ImageNet Classification with Deep Convolutional Neural Networks."

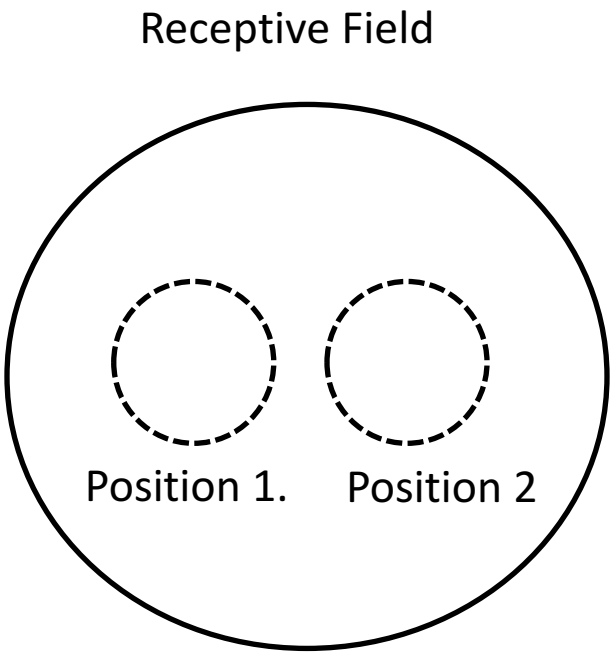
Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor
'Caffe: Convolutional Architecture for Fast Feature Embedding'

Experimental Stimuli

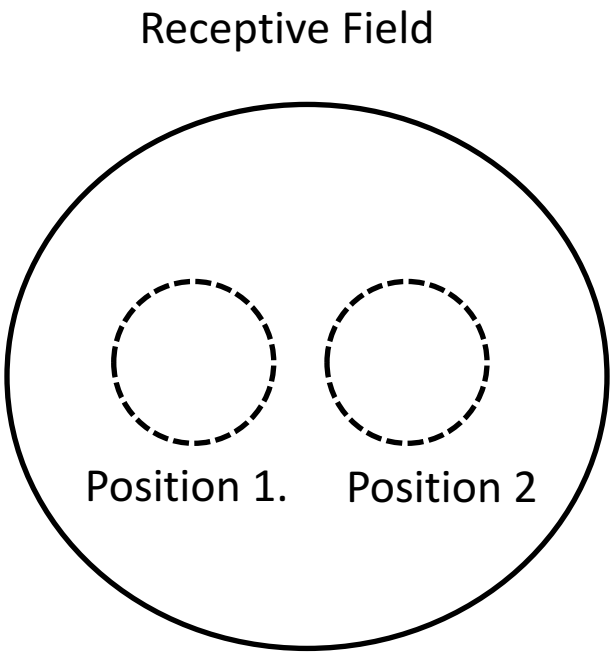


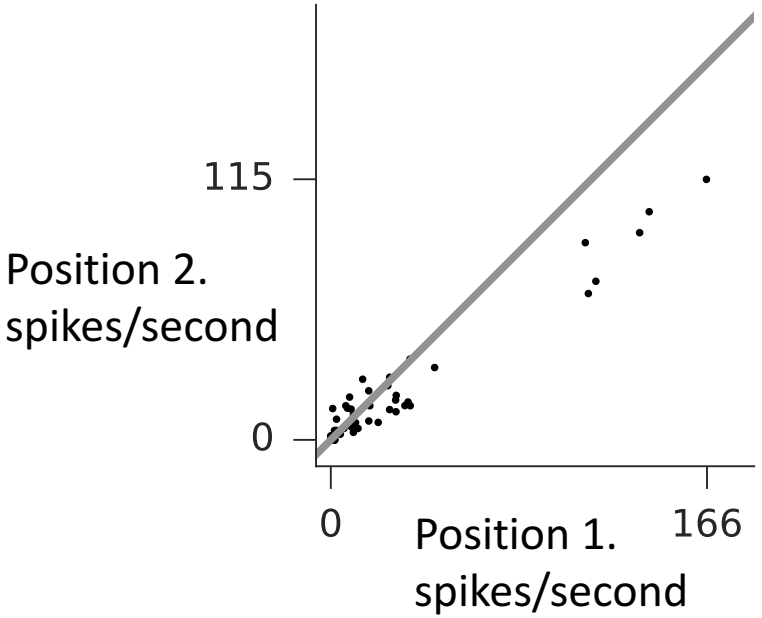
Pasupathy A, Connor CE. Shape representation in area V4: position-specific tuning for boundary conformation. J Neurophysiol. 2001

Translation Invariance



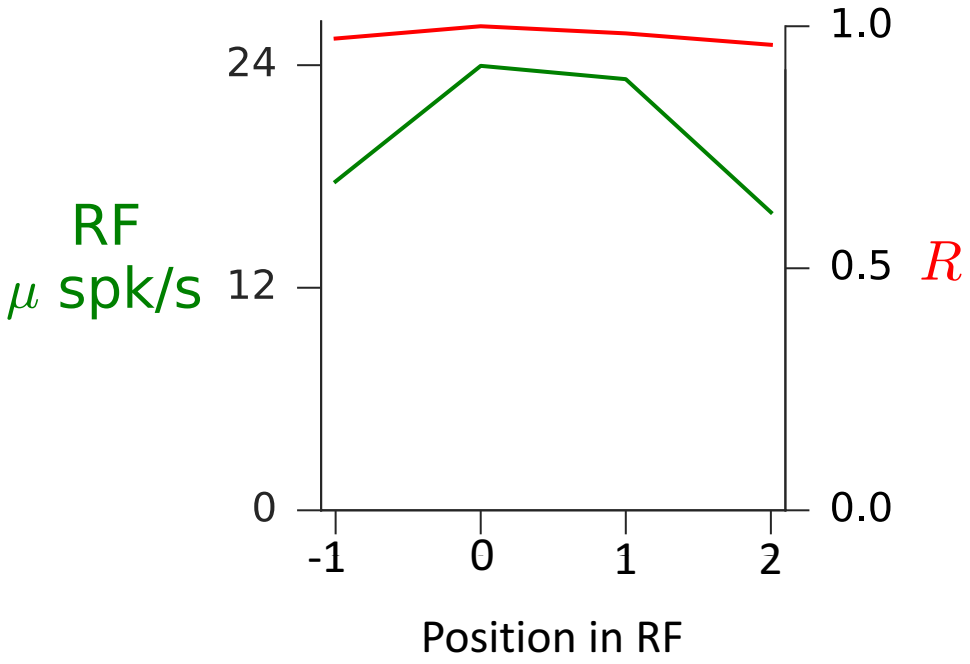
Translation Invariance





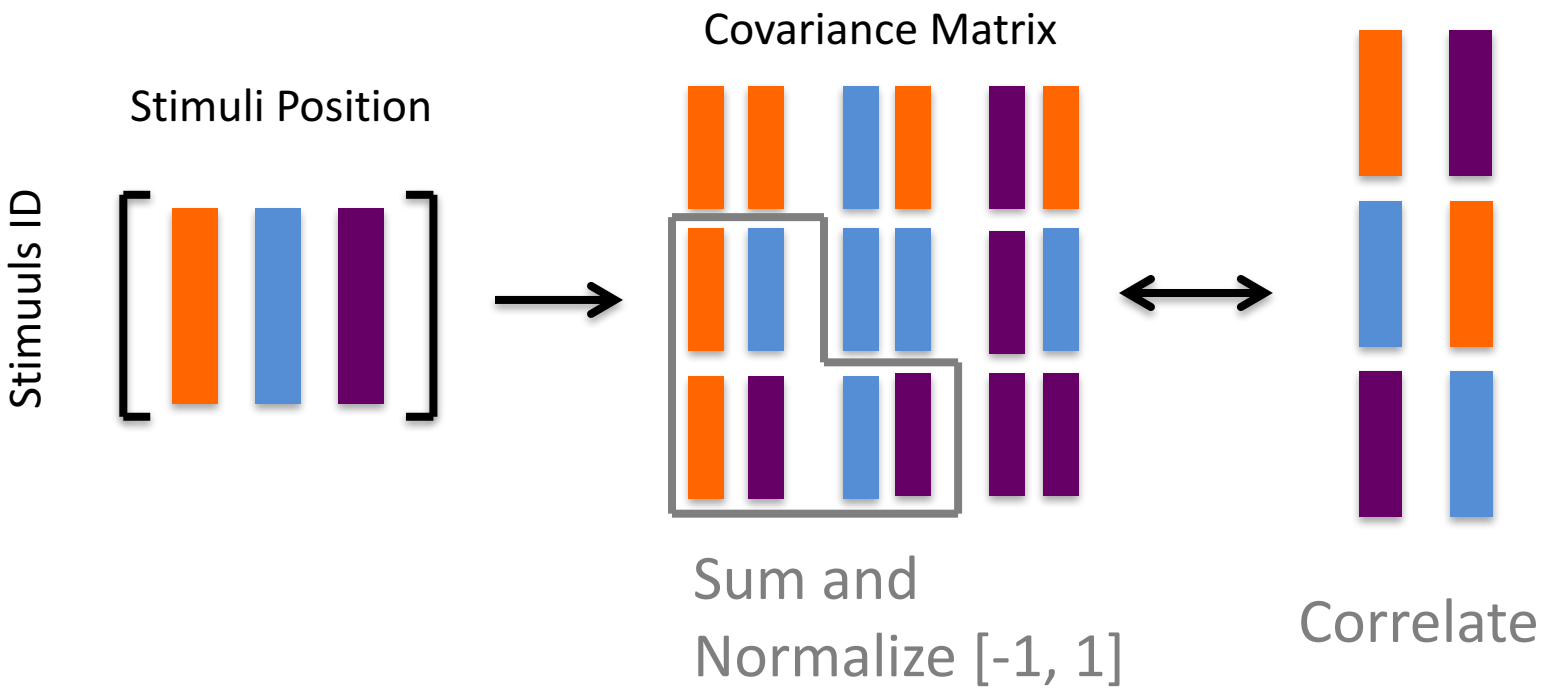
El-Shamayleh, Yasmine, and Anitha Pasupathy. “Contour Curvature As an Invariant Code for Objects in Visual Area V4.” *The Journal of Neuroscience*

V4 Translation Invariance

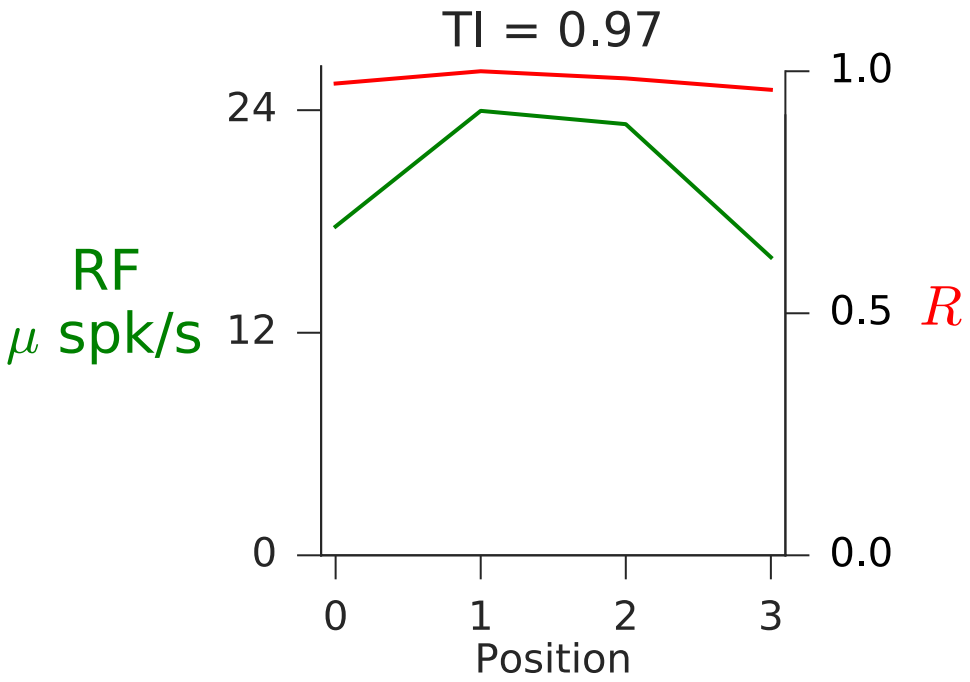


El-Shamayleh, Yasmine, and Anitha Pasupathy. “Contour Curvature As an Invariant Code for Objects in Visual Area V4.” *The Journal of Neuroscience*

Normalized sum of covariances = $$\frac{\sum_{i,j \in i \neq j} \vec{p}_i \cdot \vec{p}_j}{\sum_{i,j \in i \neq j} ||\vec{p}_i||\ ||\vec{p}_j||}$$



V4 Translation Invariance

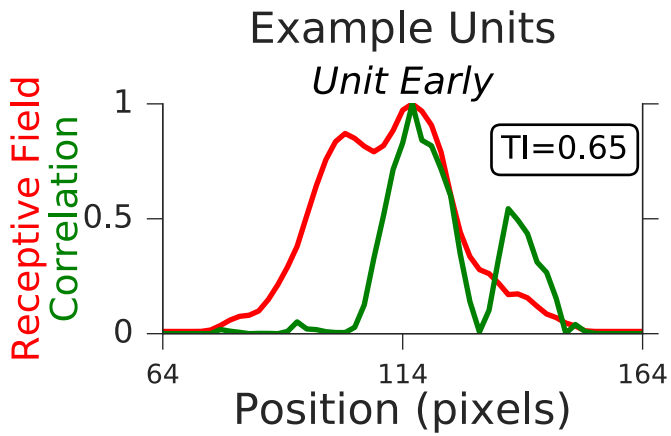


$$TI = \frac{\sum_{i,j \in i \neq j} \vec{p}_i \cdot \vec{p}_j}{\sum_{i,j \in i \neq j} ||\vec{p}_i||\ ||\vec{p}_j||}$$

El-Shamayleh, Yasmine, and Anitha Pasupathy. “Contour Curvature As an Invariant Code for Objects in Visual Area V4.” *The Journal of Neuroscience*

AlexNet Translation Invariance

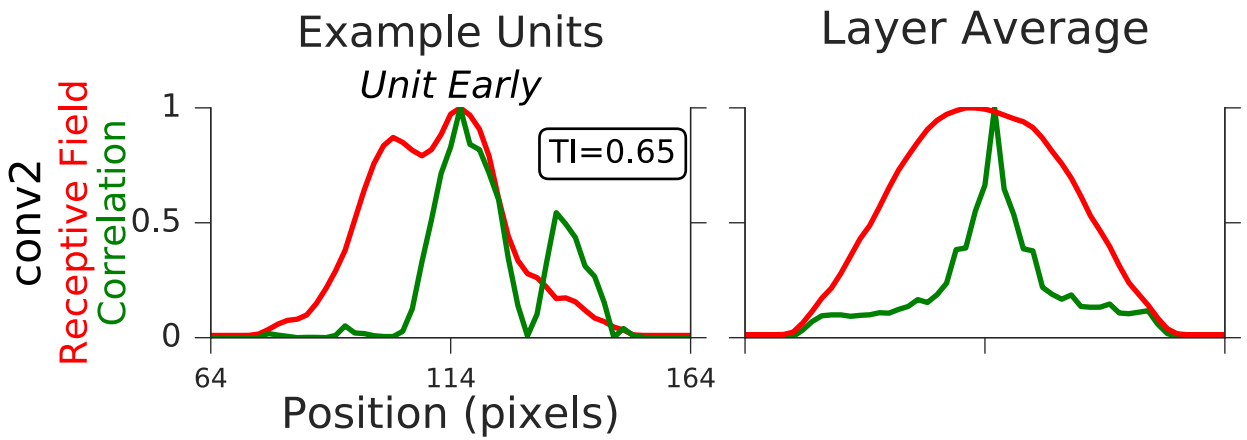
conv2



conv5

fc7

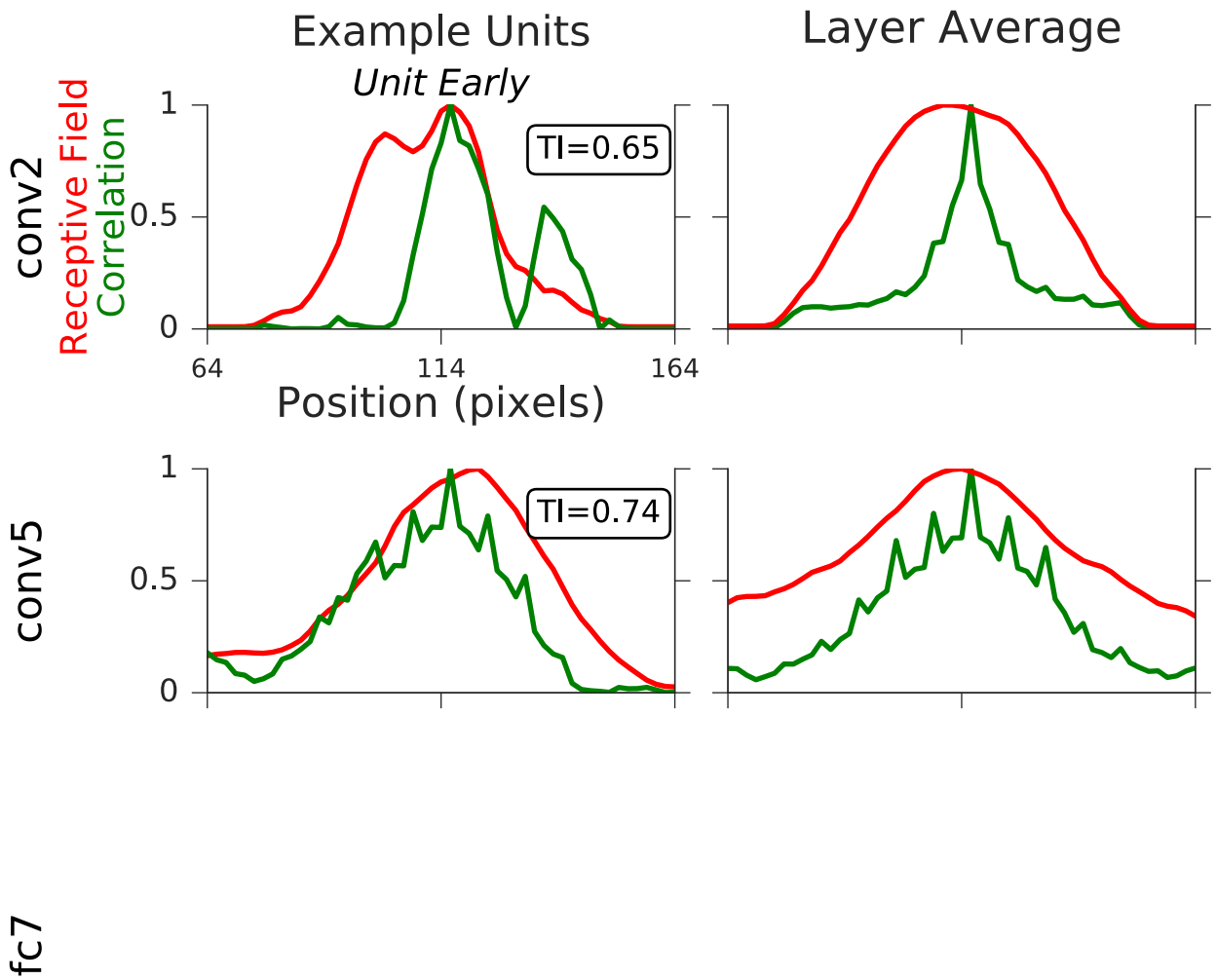
AlexNet Translation Invariance



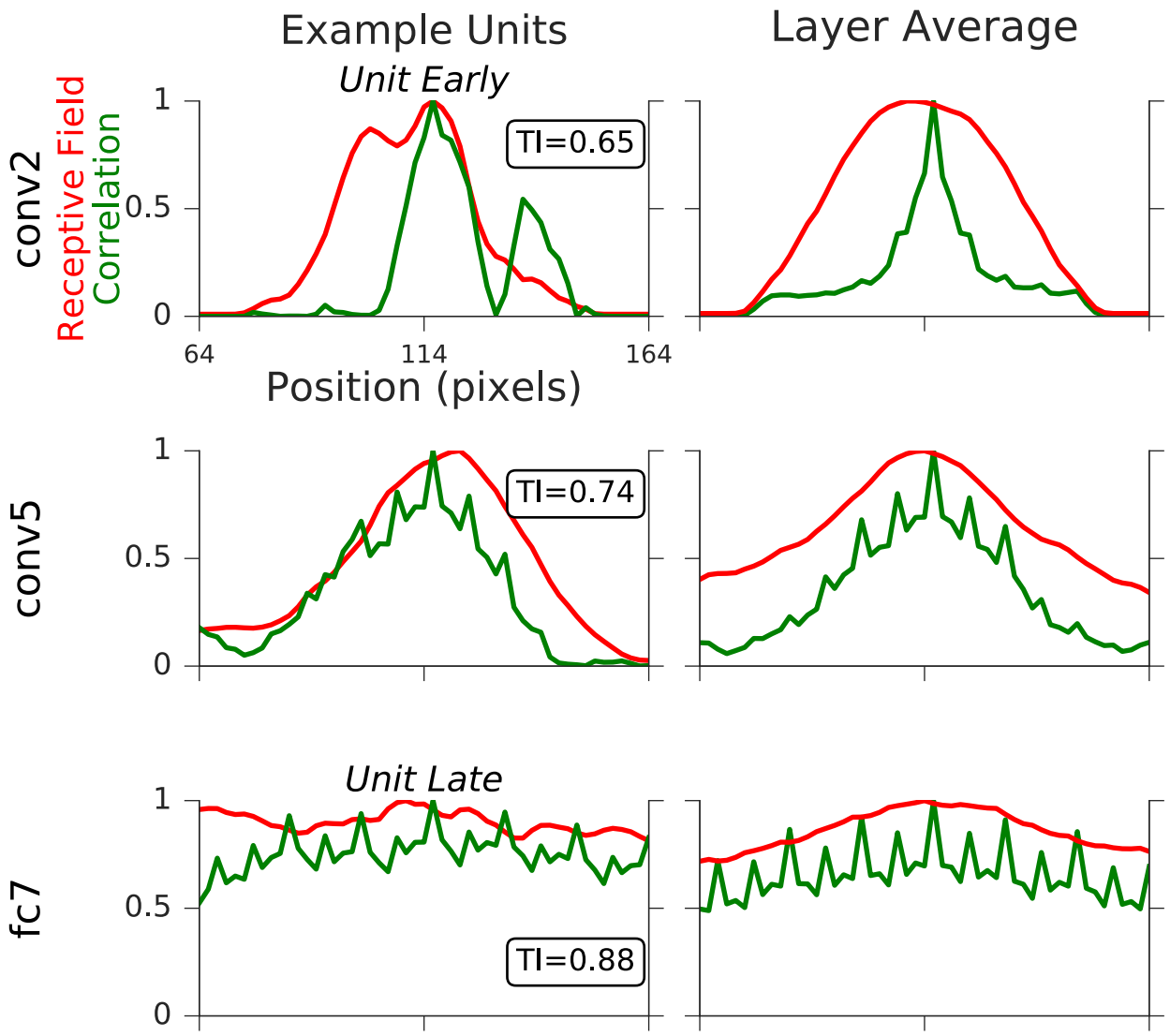
conv5

fc7

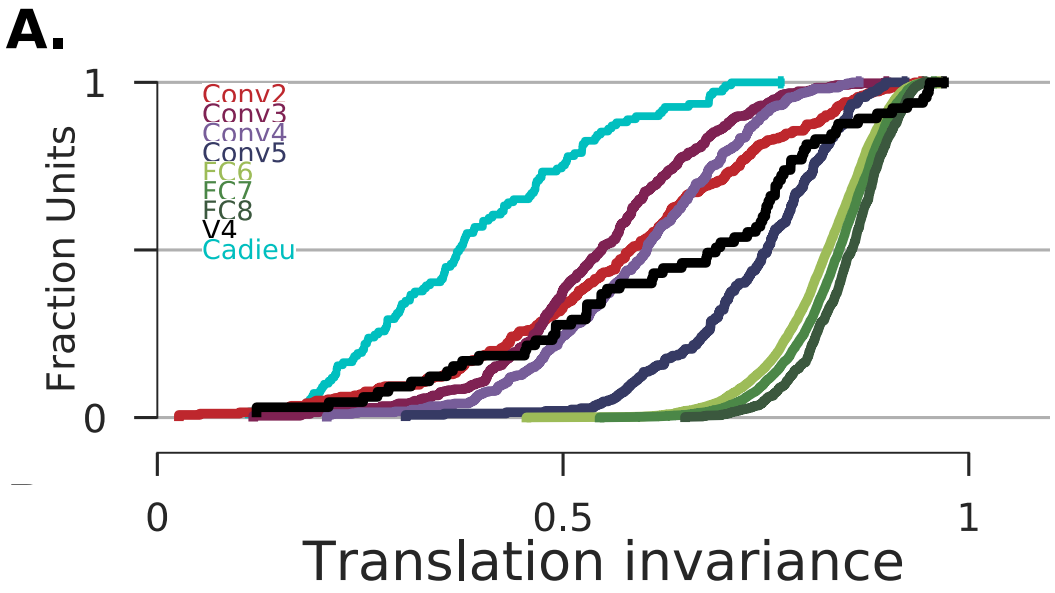
AlexNet Translation Invariance



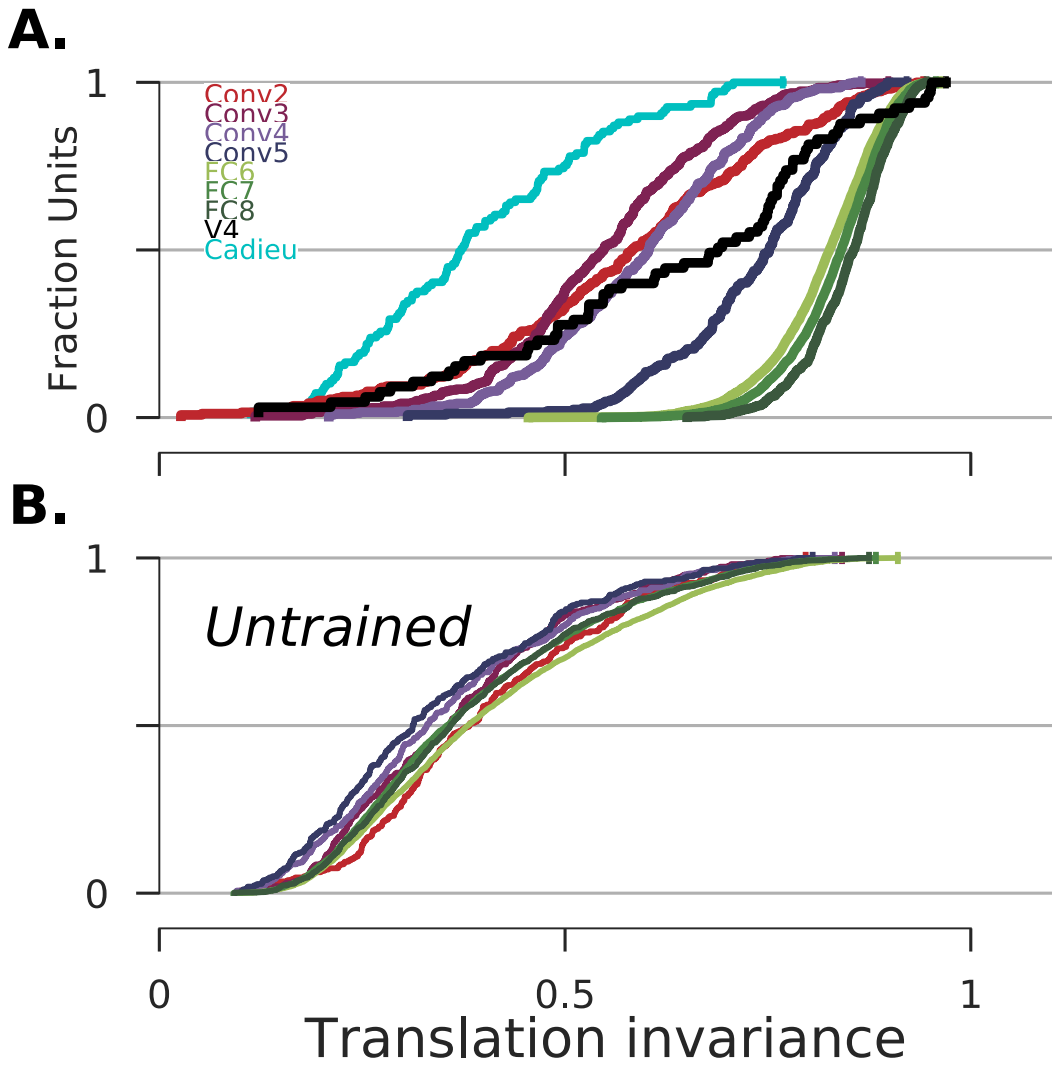
AlexNet Translation Invariance



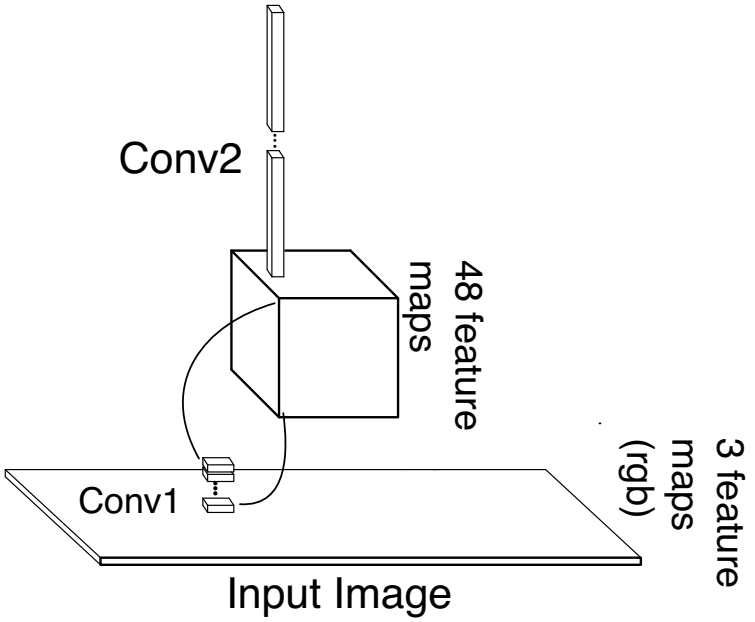
TI increases with Layer

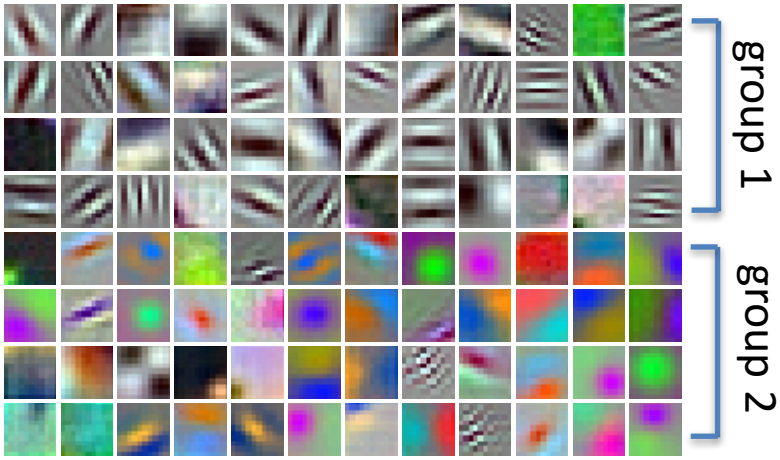


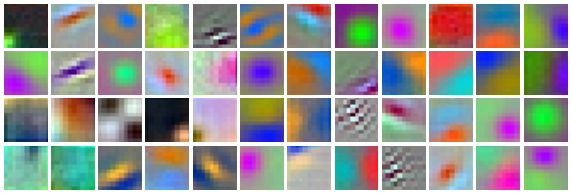
Convolution and Max-pooling don't give you translation invariance.

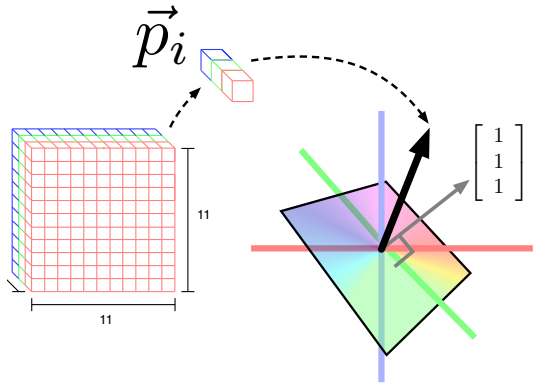


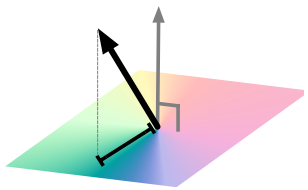
Hypothesis:
Cells with uniform spatial selectivity are translation invariant.



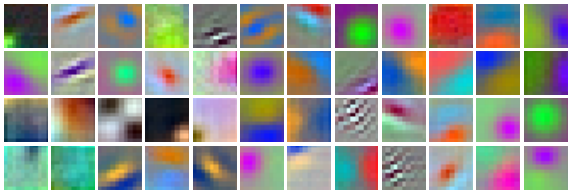


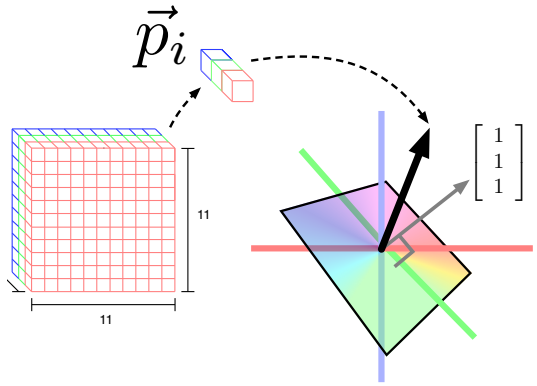


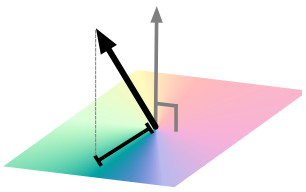




$$\frac{\sum_{i,j \in i \neq j} \vec{p}_i \cdot \vec{p}_j}{\sum_{i,j \in i \neq j} ||\vec{p}_i||\ ||\vec{p}_j||}$$

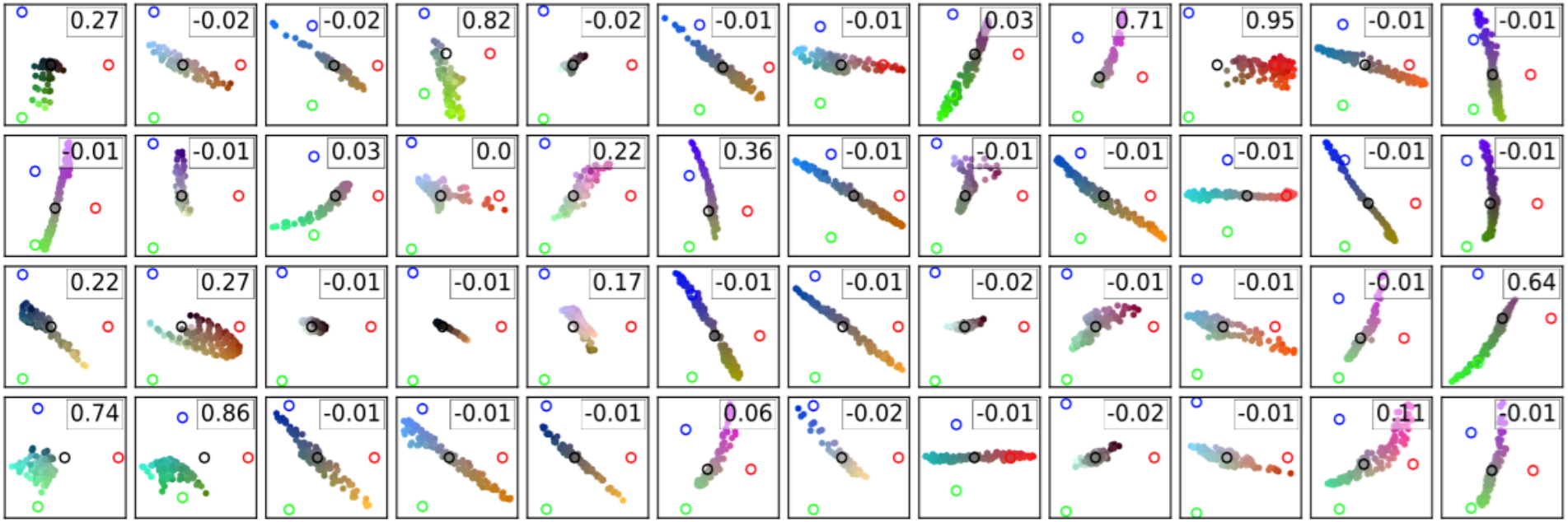




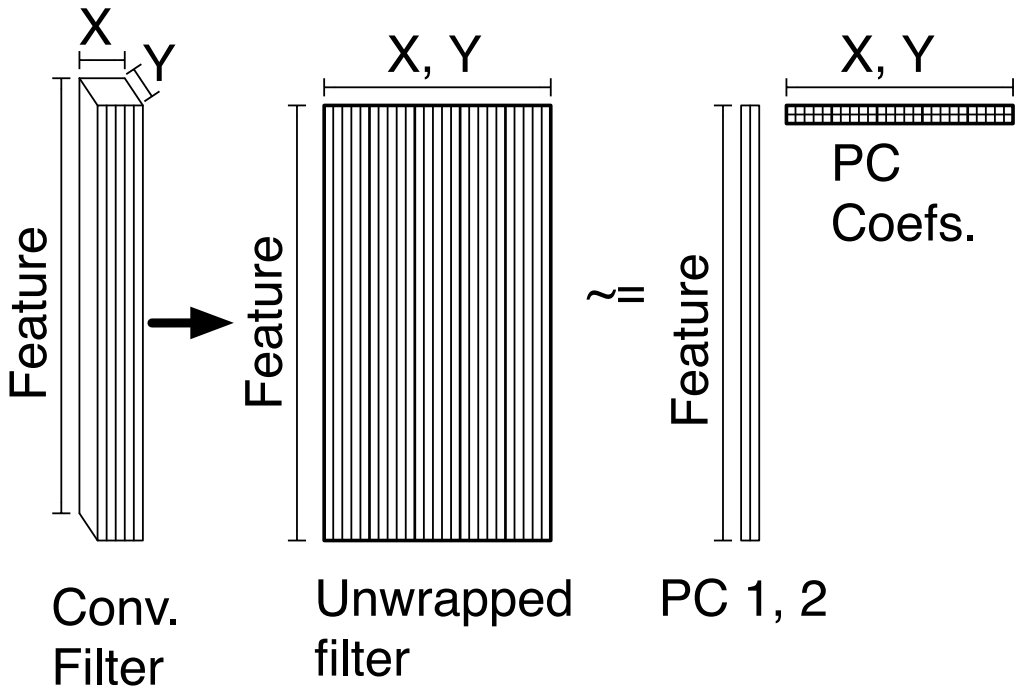


$$\frac{\sum_{i,j \in i \neq j} \vec{p}_i \cdot \vec{p}_j}{\sum_{i,j \in i \neq j} ||\vec{p}_i||\ ||\vec{p}_j||}$$

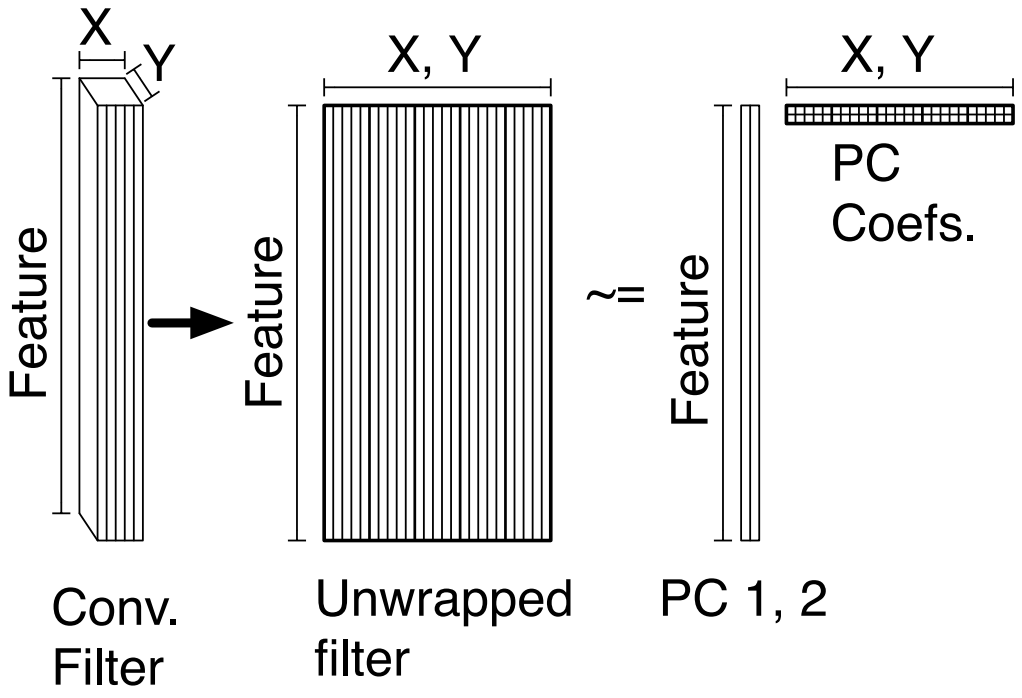
Filters plotted in the chromatic plane and their measured weight covariance.

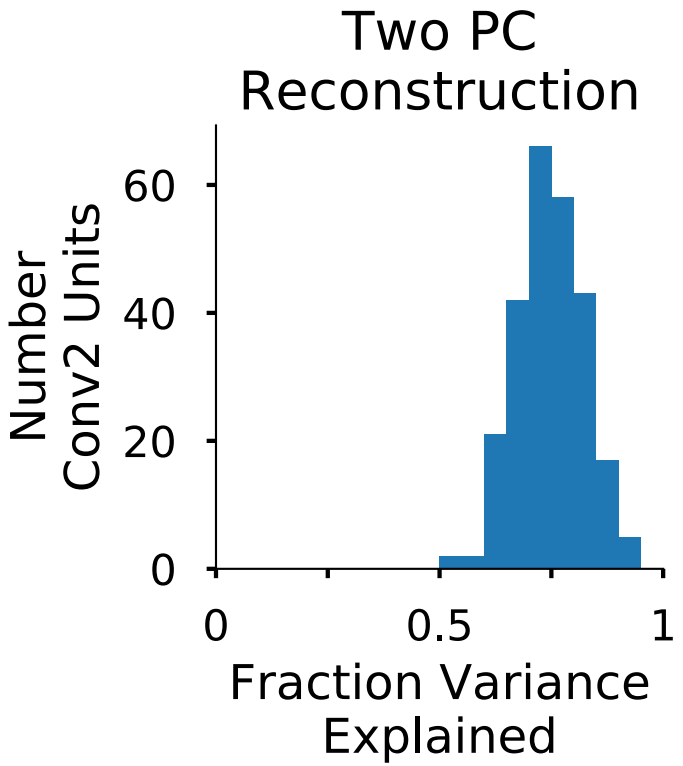


Selectivity is low-dimensional in conv2.

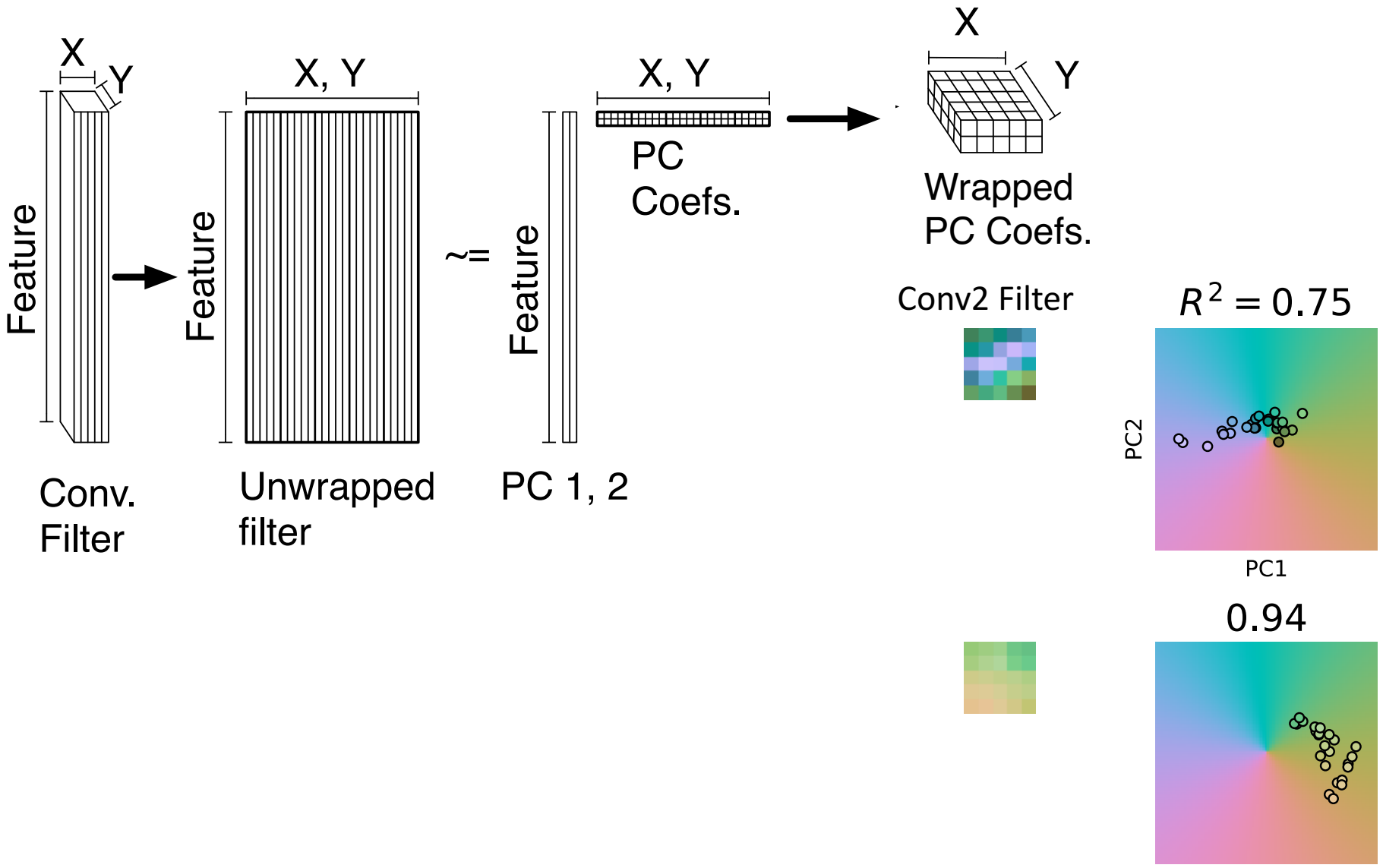


CNN selectivity is low dimensional

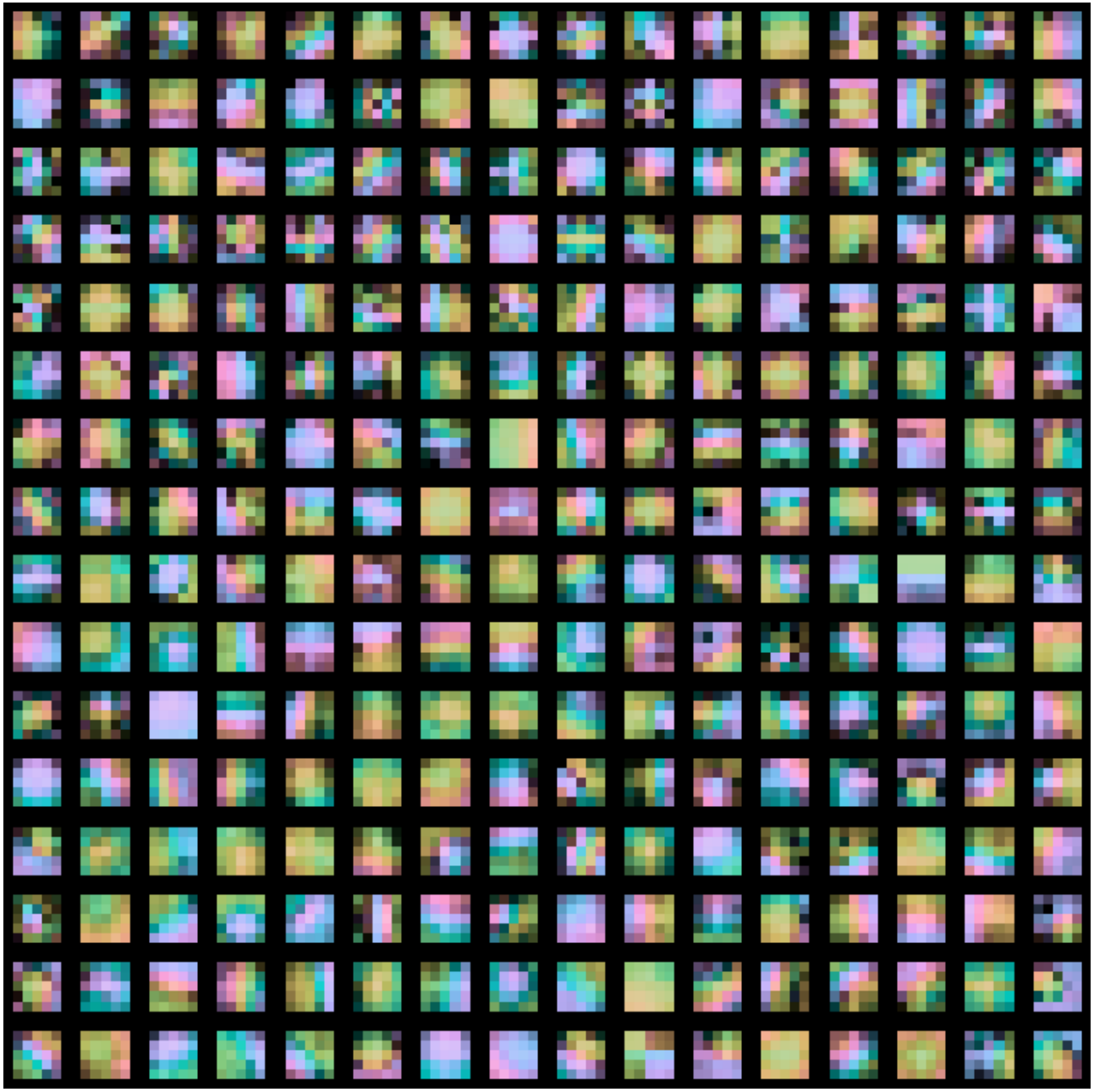




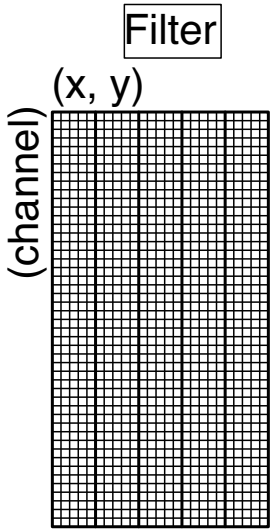
CNN selectivity is low dimensional



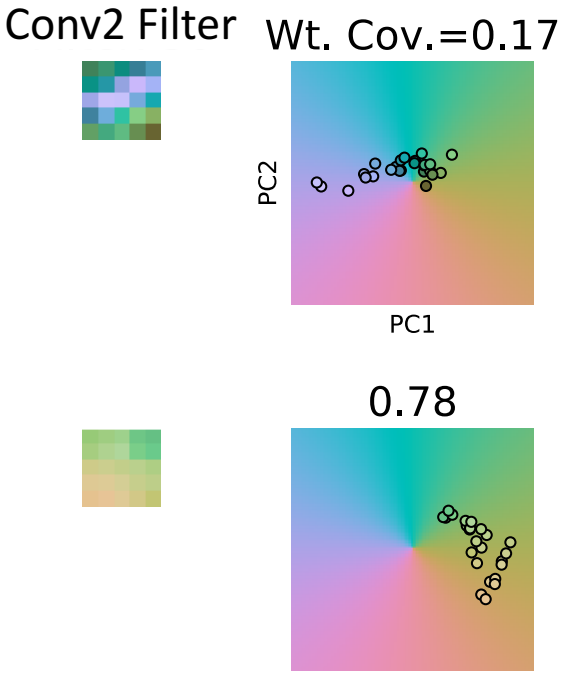
Hue=Angle(PC1 Coef., PC2 Coef.)
Luminance=Correlation(Reconstruction, Original)



Weight Covariance

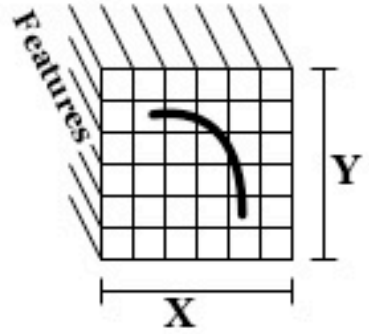


$$\frac{\sum_{i,j \in i \neq j} \vec{p}_i \cdot \vec{p}_j}{\sum_{i,j \in i \neq j} ||\vec{p}_i||\ ||\vec{p}_j||}$$



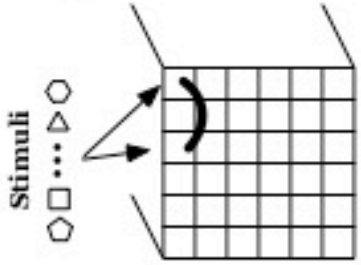
Linear models work in hidden layer.

Weight correlation

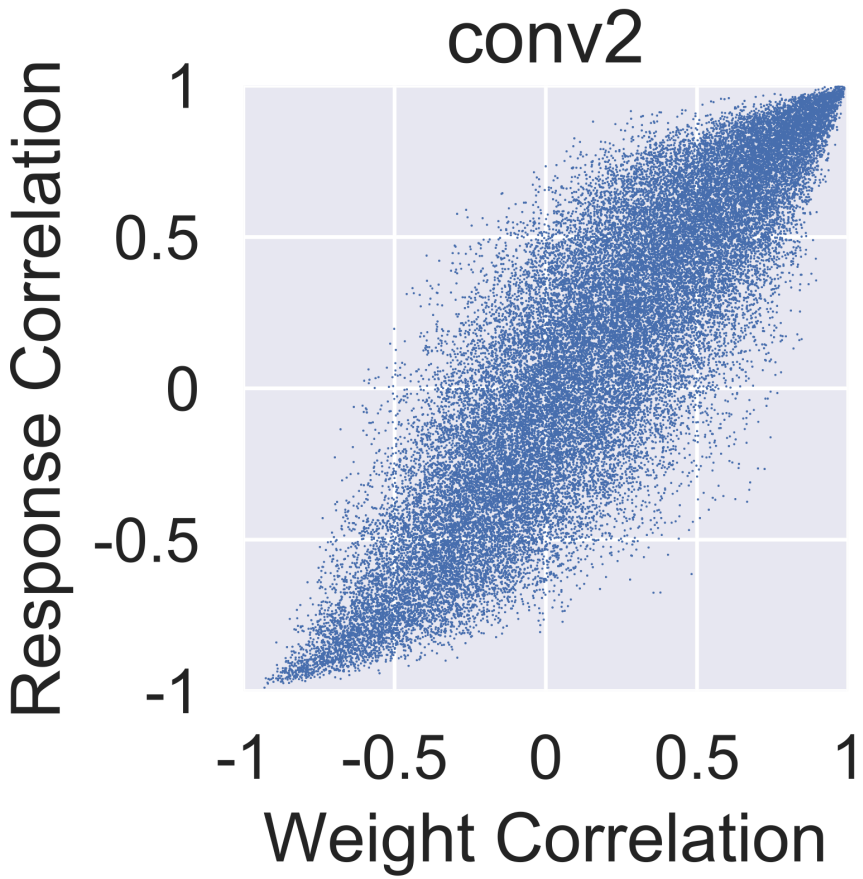


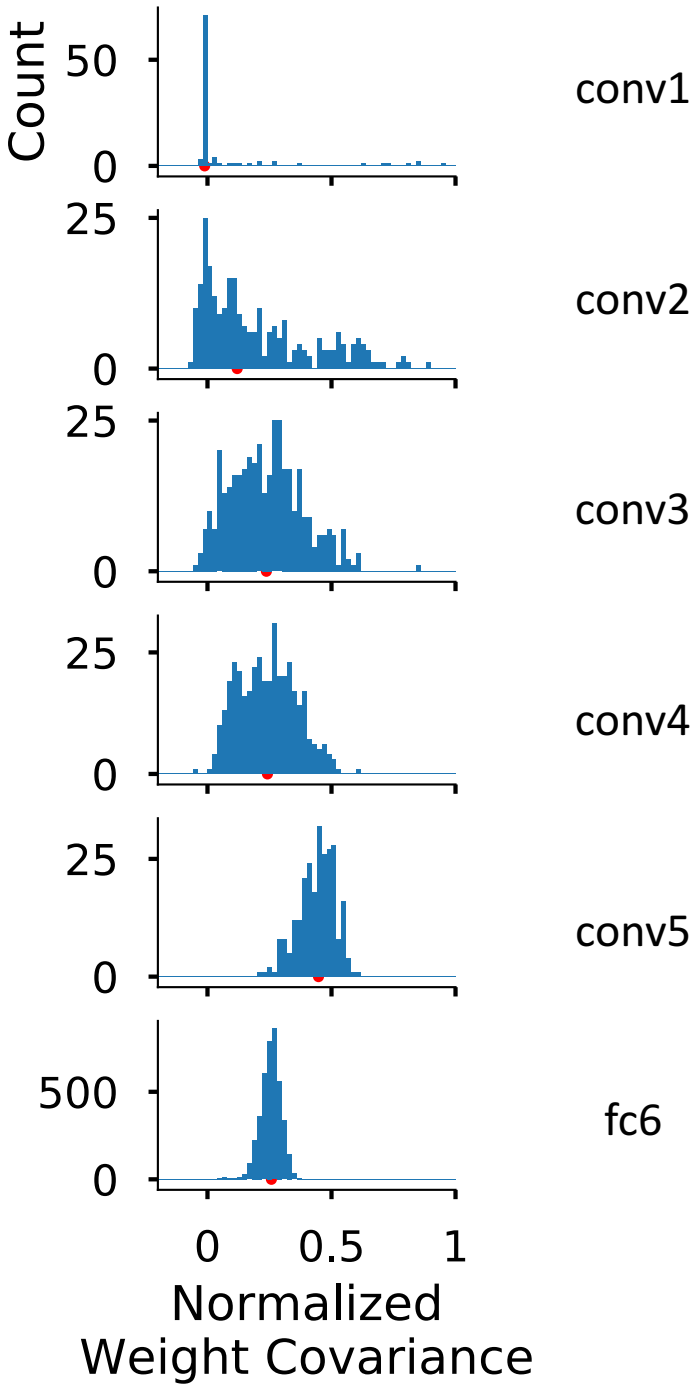
Get correlation between spatial channel's feature weights.

Response Correlation

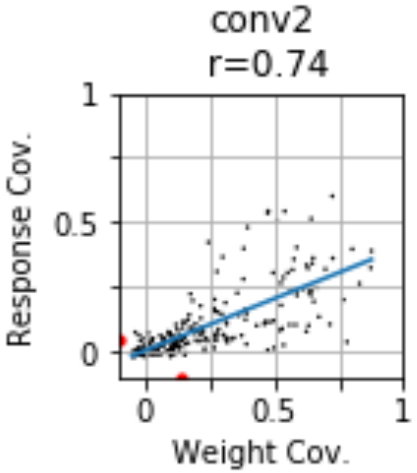


Stimulate all spatial channels get correlations between responses.

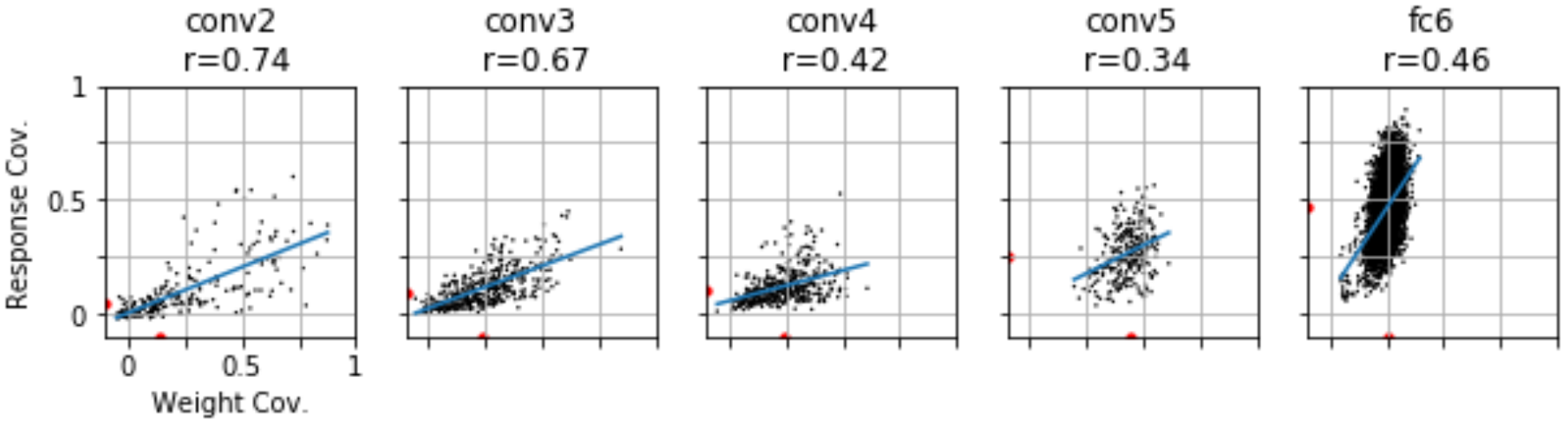


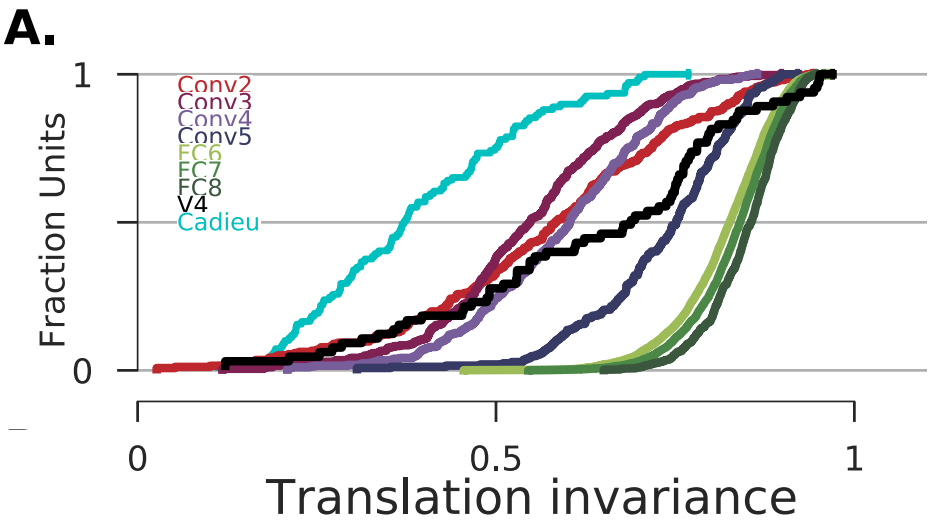

Count
50
0
conv1
25
0
conv2
25
0
conv3
25
0
conv4
25
0
conv5
500
0
fc6
0
0.5
1
Normalized
Weight Covariance

Weight Covariance Predicts TI

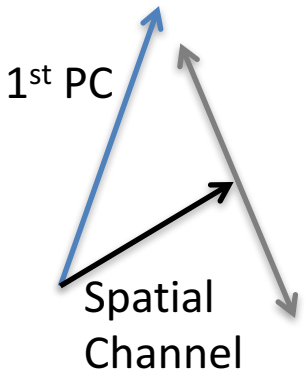


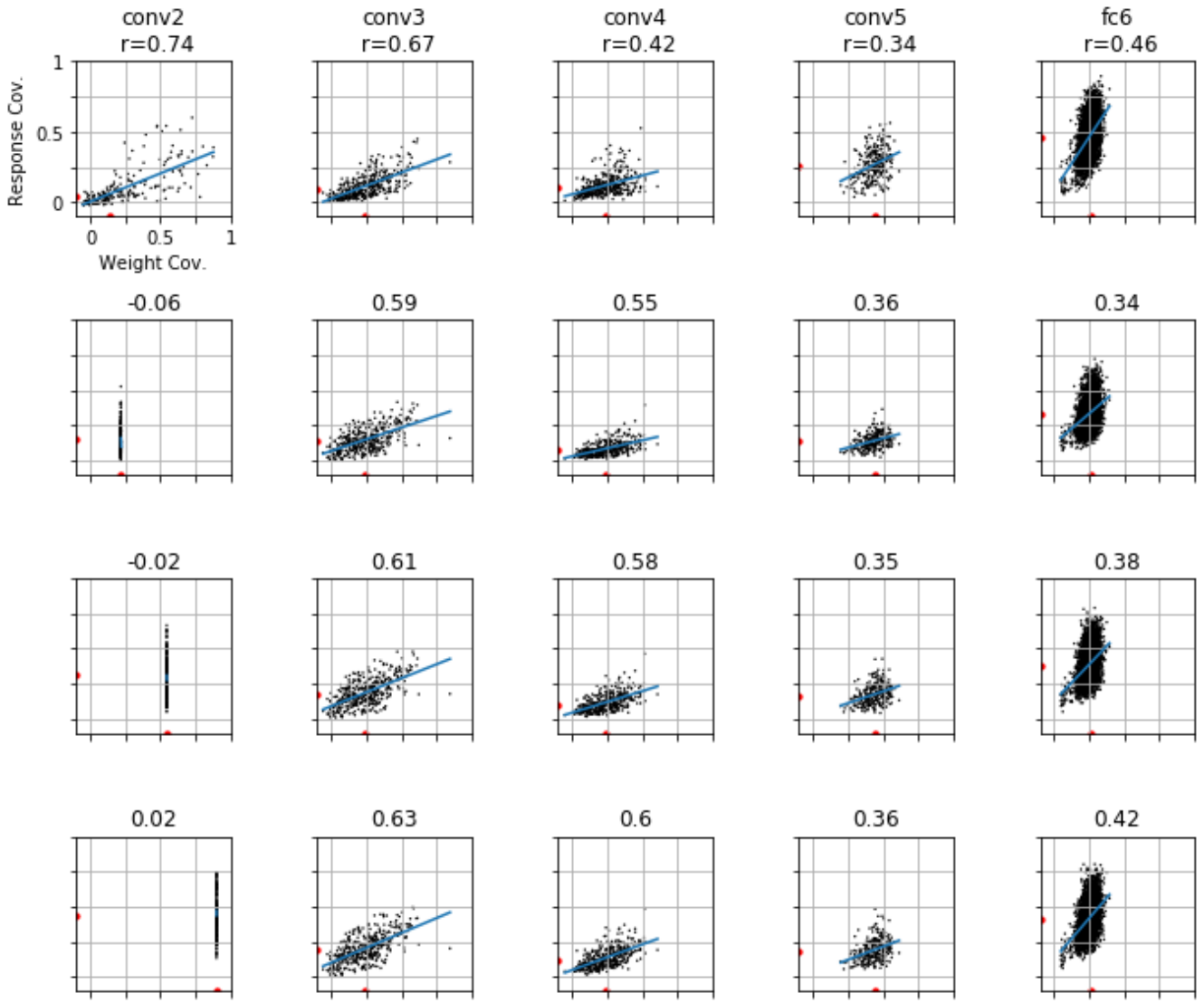
Weight Covariance Predicts Translation Invariance



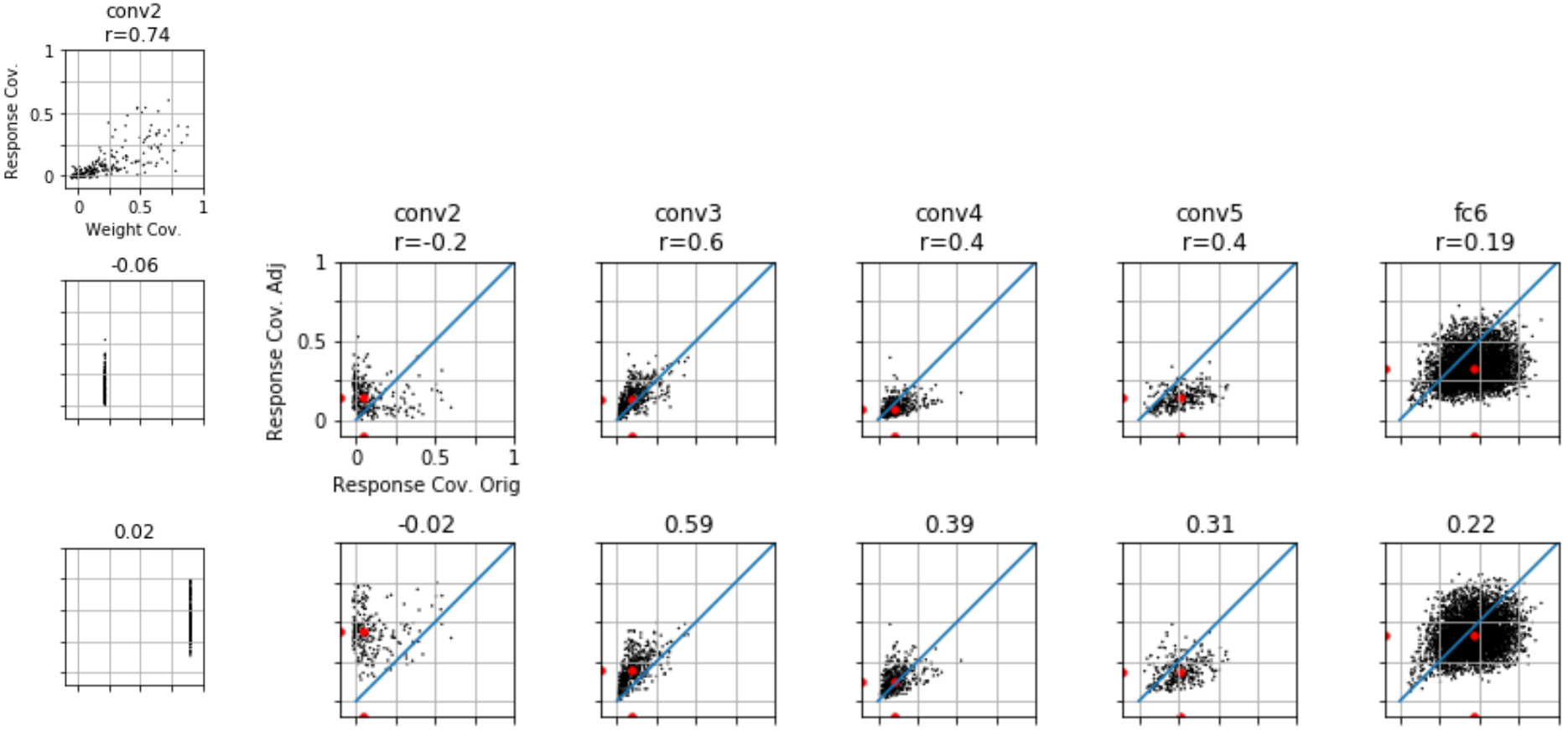


Weight Covariance Increases Translation Invariance

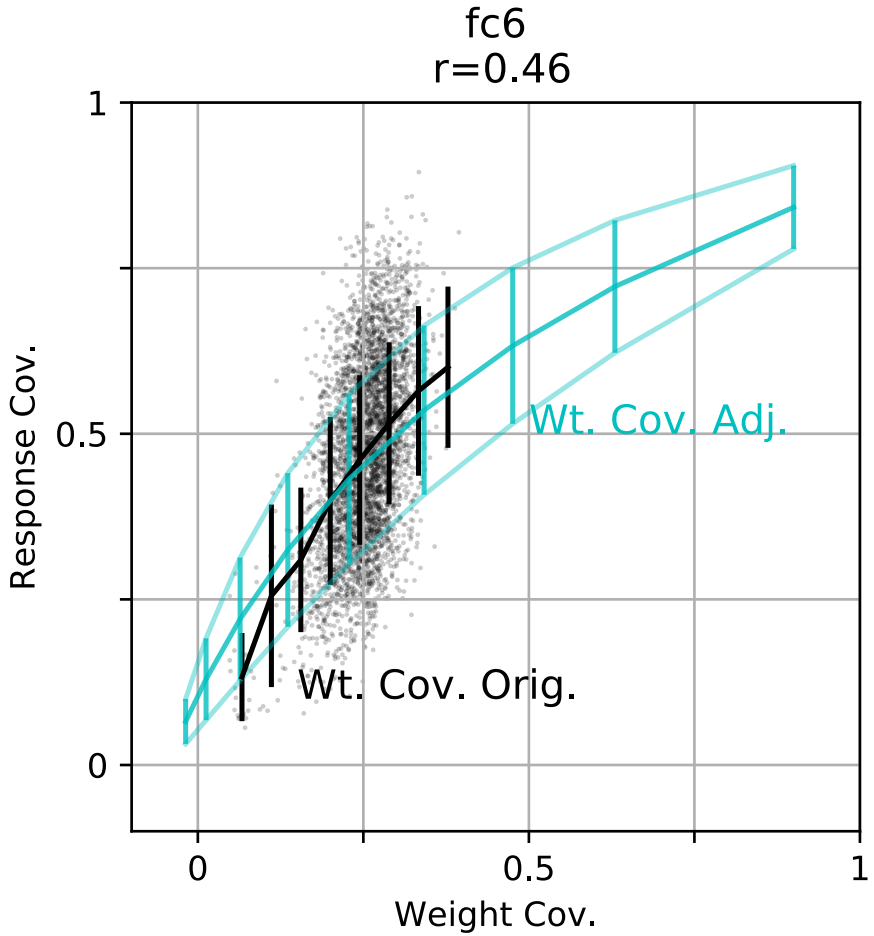


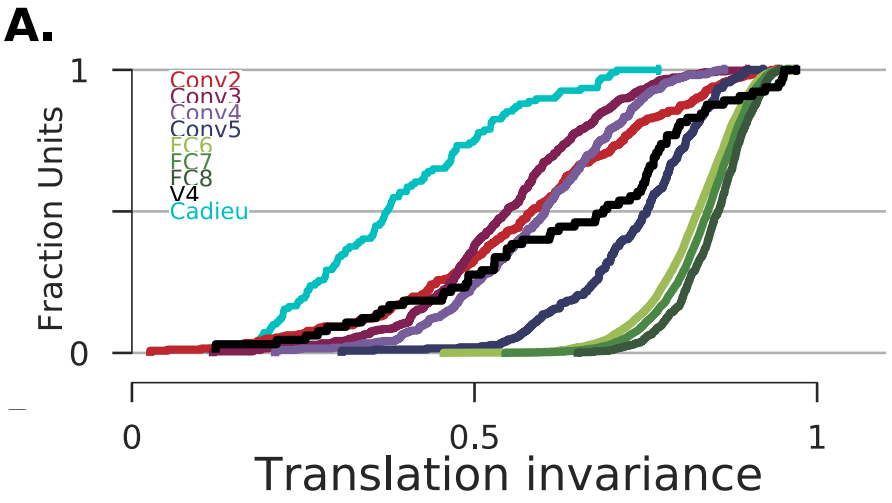


Weight Covariance Increases Translation Invariance



TI Could Have Been Higher





Thanks! Questions?

Weight Covariance Predicts Translation Invariance