

It allows parameters learned for one task to be shared even for a different task (transfer learning).

It reduces the total number of parameters, thus reducing overfitting.

It allows gradient descent to set many of the parameters to zero, thus making the connections sparse.

It allows a feature detector to be used in multiple locations throughout the whole input image/input volume.

10. In lecture we talked about "sparsity of connections" as a benefit of using convolutional layers. What does this mean?

Each filter is connected to every channel in the previous layer.

Each activation in the next layer depends on only a small number of activations from the previous layer.

Each layer in a convolutional network is connected only to two other layers

Regularization causes gradient descent to set many of the parameters to zero.

© 2020 GitHub, Inc. Terms Privacy Security Status Help 🕞 Contact GitHub Pricing API Training Blog Abo