Pratik Pokharel(50539742)

Paper Summary

The experiments described challenge the popular belief that convolutional neural networks (CNNs) are strictly invariant to translation. They provide insights into why and when convolution is or is not strictly invariant to translation:

1. Convolution and Limited Built-In Invariance:

Convolution is the fundamental operation in CNNs, and it does possess a degree of translational invariance. It allows networks to detect features and patterns at different locations within an image, which is an essential part of achieving translation invariance. Convolution does provide some level of built-in translational invariance, but this invariance is not necessarily strict or complete. When common CNN networks were trained on objects at one location and tested on novel locations, they demonstrated a significant and consistent drop in performance. This means that these networks(like ResNet-50 and others), do not possess inherent architectural invariance. The extent of invariance achieved through convolution alone is limited and may not be sufficient for recognizing objects across a wide range of translations.

2. Role of Pre-training:

- A notable finding was that pre-training CNNs on fully translated datasets, regardless of their complexity (e.g., from ImageNet to MNIST), greatly improved their ability to achieve translation invariance. This improvement was evident in the similarity of hidden unit activations when recognizing translated objects. Pre-training imparts the network with the capability to recognize objects at novel locations, even if the network architecture lacks inherent architectural invariance.

- **Importance of Data Variation:** Pre-training is more effective when the network is exposed to fully-translated objects during training. Data variation, such as translating objects across the canvas, is highly influential for improving translation invariance.

3. Influence of Fine-Tuning and Interference:

Fine-tuning on one-location datasets can introduce interference, potentially leading to a loss of translation invariance. This suggests that fine-tuning may disrupt the previously acquired invariance to some extent.

4. Role of Architectural Modifications:

The experiments included some architectural modifications, such as Global Average Pooling, Anti-Aliasing, and Fully Convolutional operations, aimed at enhancing online invariance. However, these modifications showed only marginal improvements in performance and, in some cases, resulted in poor performance.

5. Complexity of Training Data:

The complexity of the training data, such as fully-translated datasets, can impact translation invariance. More complex datasets may lead to reduced translation invariance in certain cases.

6. Generalization and Learning Invariances:

- **Learning Invariances:** The experiments highlight that CNNs can learn translation invariance and other invariances from their training data, even when their architecture does not inherently support these invariances. This suggests that CNNs have the capacity to extract perceptual regularities from their environment through training.

- **Generalization Challenges:** CNNs often struggle with generalization, especially in novel environments. While they can learn specific perceptual regularities, they may face challenges when combinatorial generalization(the ability to understand and produce novel combinations of already familiar elements) is required.

In summary, while convolution provides a foundation for achieving translation invariance, the strictness of this invariance depends on various factors, including pre-training, data variation, fine-tuning, and the network's architecture. The experiments challenge the notion of strict architectural invariance and emphasize the importance of careful training strategies and the role of training data in achieving robust translation invariance in CNNs.

References:

[1] Valerio Biscione, and Jeffrey Bowers. Convolutional Neural Networks Are Not Invariant to Translation, but They Can Learn to Be.

[2] Ryan Blything, Valerio Biscione, and Jeffrey Bowers. A case for robust translation tolerance in humans and CNNs. A commentary on Han et al.

[3] Christina M Funke, Judy Borowski, Karolina Stosio, Wieland Brendel, Thomas S A Wallis, and Matthias Bethge. The Notorious Difficulty of Comparing Human and Machine Perception.

[4] K. Ko ka. Principles of gestalt psychology. Taylor and Francis, 1 2013.