

Detail Publikasi

Judul Artikel : Visualizing and Understanding Convolutional Networks
Penulis : Matthew D. Zeiler, Rob Fergus
Tahun Terbit : 2013
Link Artikel :]<https://arxiv.org/abs/1311.2901>

Latar Belakang

Pada tahun 2013, Convolutional Neural Networks (CNNs) telah menunjukkan performa yang luar biasa dalam berbagai tugas Computer Vision, termasuk klasifikasi gambar dan deteksi objek. Namun, banyak peneliti dan praktisi merasa kesulitan memahami apa yang sebenarnya dipelajari oleh CNN, terutama di lapisan-lapisan yang lebih dalam. Pemahaman yang terbatas tentang cara kerja internal CNN membuat pengembangan dan peningkatan arsitektur jaringan menjadi tantangan. Oleh karena itu, penelitian ini berfokus pada pengembangan metode visualisasi untuk membantu memahami apa yang terjadi di dalam jaringan selama proses pembelajaran dan pengenalan pola.

Isi Artikel

Artikel ini mengeksplorasi teknik visualisasi yang baru untuk membantu menjelaskan bagaimana CNN mempelajari representasi hierarkis dari data input, khususnya gambar. Poin-poin utama yang dibahas meliputi:

- Penjelasan tentang Deconvolutional Network (DeconvNet), metode visualisasi yang digunakan untuk memeriksa aktivasi di setiap lapisan CNN.
- Eksperimen dengan berbagai jaringan untuk menunjukkan bahwa jaringan yang lebih dalam secara signifikan meningkatkan kemampuan pengenalan pola dibandingkan dengan jaringan yang lebih dangkal.
- Pemahaman tentang apa yang dipelajari oleh setiap lapisan jaringan CNN, dari fitur-fitur sederhana seperti tepi dan sudut pada lapisan awal hingga representasi kompleks seperti objek lengkap pada lapisan yang lebih dalam.

Dataset

Penelitian ini menggunakan dataset ImageNet, salah satu dataset yang paling umum digunakan dalam penelitian Computer Vision. ImageNet berisi jutaan gambar yang dikategorikan ke dalam ribuan kelas objek yang berbeda. AlexNet, arsitektur CNN yang

dilatih menggunakan dataset ImageNet, juga menjadi fokus utama eksperimen dalam artikel ini. Dataset ini sangat relevan karena:

- Mengandung variasi besar dalam bentuk, ukuran, dan posisi objek, yang memerlukan jaringan yang kuat untuk menghasilkan hasil klasifikasi yang akurat.
- Digunakan dalam berbagai kompetisi seperti ImageNet Large Scale Visual Recognition Challenge (ILSVRC) yang memacu pengembangan model-model CNN baru yang lebih efektif.

Metodologi

Metodologi yang digunakan berfokus pada penerapan teknik visualisasi **Deconvolutional Network (DeconvNet)** untuk memahami CNN. Berikut adalah langkah-langkah utama dalam metodologi:

- **Arsitektur DeconvNet:** Teknik ini membalik proses konvolusi dan pooling yang terjadi di dalam CNN untuk melihat apa yang diaktifkan oleh setiap lapisan dari input gambar asli. DeconvNet membongkar data yang dihasilkan oleh CNN untuk memetakan kembali aktivasi di setiap lapisan ke domain gambar.
- **Visualisasi Filter:** Dengan menggunakan DeconvNet, peneliti dapat memvisualisasikan fitur yang dipelajari di setiap lapisan CNN. Lapisan awal biasanya mendeteksi fitur sederhana seperti tepi dan tekstur, sementara lapisan yang lebih dalam mendeteksi bagian-bagian objek yang lebih kompleks.
- **Eksperimen dengan Arsitektur Jaringan yang Berbeda:** Peneliti membandingkan kinerja dan visualisasi jaringan dengan kedalaman yang berbeda, termasuk **AlexNet**, untuk melihat bagaimana jaringan yang lebih dalam mempelajari representasi yang lebih kompleks dan kuat.

Hasil dan Pembahasan

- **Visualisasi Aktivasi Lapisan:** Hasil eksperimen menunjukkan bahwa lapisan awal CNN mempelajari fitur-fitur dasar seperti tepi, sudut, dan tekstur, yang bersifat lebih umum dan bisa digunakan untuk objek apapun. Lapisan yang lebih dalam di CNN menangkap komponen objek yang lebih kompleks seperti bagian tubuh, wajah, atau bahkan keseluruhan objek.
- **Pentingnya Kedalaman Jaringan:** Eksperimen menunjukkan bahwa jaringan yang lebih dalam (seperti AlexNet) mampu mempelajari pola yang lebih kompleks dan spesifik, yang memungkinkan jaringan tersebut unggul dalam tugas klasifikasi gambar. Semakin dalam jaringan, semakin banyak informasi yang dapat diproses dan semakin akurat prediksi yang dihasilkan.
- **Interpretasi Visualisasi:** Visualisasi dari DeconvNet membantu para peneliti melihat bagaimana CNN merespons gambar input tertentu, yang memungkinkan untuk memahami representasi hierarkis yang dipelajari oleh jaringan. Ini juga menunjukkan

bahwa lapisan pooling berperan penting dalam menjaga invariansi terhadap pergeseran posisi objek dalam gambar.

- **Perbandingan dengan Jaringan yang Dangkal:** Jaringan yang lebih dangkal, meskipun lebih mudah dipahami, tidak mampu menghasilkan representasi yang kuat, sehingga performa klasifikasinya lebih buruk dibandingkan dengan jaringan yang lebih dalam.

Kesimpulan

- Teknik Visualisasi sebagai Alat Penting: Teknik visualisasi DeconvNet yang dikembangkan dalam penelitian ini memberikan pemahaman yang lebih dalam tentang bagaimana CNN bekerja. Visualisasi ini mempermudah pemahaman tentang bagaimana jaringan memproses dan mempelajari pola-pola yang berbeda di setiap lapisan.
- Kedalaman CNN Penting untuk Kinerja Optimal: Jaringan yang lebih dalam terbukti lebih efektif dalam tugas pengenalan gambar, karena mampu menangkap lebih banyak pola yang kompleks dan variatif. Semakin dalam jaringan, semakin baik dalam menangani tugas-tugas yang lebih rumit.
- Kontribusi untuk Pengembangan CNN Masa Depan: Pemahaman yang lebih baik tentang apa yang terjadi di dalam jaringan CNN dapat membantu para peneliti merancang arsitektur yang lebih efisien dan efektif. Penelitian ini membuka pintu untuk eksplorasi lebih lanjut dalam meningkatkan kinerja jaringan dengan lebih baik memahami proses internalnya.

Secara keseluruhan, artikel ini memberikan kontribusi besar dalam meningkatkan pemahaman tentang bagaimana CNN mempelajari dan mengenali pola, dan teknik visualisasi yang diperkenalkan memainkan peran penting dalam proses ini.