

Abstrak

Pengembangan kecerdasan buatan (AI) dalam bidang computer vision menghadirkan tantangan kompleks yang mencakup optimisasi model untuk deployment, pengelolaan bias dan etika, serta penerapan metode pembelajaran canggih seperti Zero-Shot Learning (ZSL). Laporan ini menganalisis berbagai aspek tersebut berdasarkan materi yang diberikan, dengan fokus pada studi kasus ImageNet Roulette dan upaya komunitas seperti Hugging Face dalam menciptakan AI yang etis dan inklusif. Melalui pendekatan multidimensional, laporan ini menawarkan wawasan mendalam mengenai strategi optimisasi, mitigasi bias, serta penerapan prinsip-prinsip etika dalam pengembangan model AI.

I. Pendahuluan

Dalam era digital saat ini, model AI, khususnya dalam computer vision, memainkan peran vital dalam berbagai aplikasi seperti kendaraan otonom, augmented reality, dan pengenalan wajah. Namun, pengembangan AI yang efektif tidak hanya bergantung pada akurasi model, tetapi juga mencakup optimisasi untuk perangkat terbatas, pengelolaan bias, serta penerapan etika yang kuat. Laporan ini bertujuan untuk menganalisis dan memahami berbagai aspek tersebut, dengan fokus pada studi kasus ImageNet Roulette dan inisiatif Hugging Face dalam menciptakan AI yang adil dan inklusif.

II. Optimisasi Model untuk Deployment

A. Pengenalan

Optimisasi model adalah proses modifikasi model AI untuk meningkatkan efisiensi dalam hal penggunaan sumber daya, kecepatan inferensi, dan ukuran model. Langkah ini krusial karena perangkat yang digunakan untuk deployment sering kali memiliki spesifikasi lebih rendah dibandingkan perangkat yang digunakan untuk pelatihan model.

B. Definisi Model Optimization

Model optimization melibatkan teknik-teknik seperti pruning, quantization, knowledge distillation, low-rank approximation, dan kompresi model spesifik hardware. Teknik-teknik ini membantu mengurangi kompleksitas dan ukuran model, meningkatkan kecepatan inferensi, serta menyesuaikan model dengan perangkat keras target.

C. Pentingnya Optimisasi dalam Computer Vision

Optimisasi model sangat penting untuk:

1. **Keterbatasan Sumber Daya:** Mengurangi kebutuhan memori dan daya komputasi pada perangkat terbatas seperti smartphone dan IoT.
2. **Kebutuhan Latensi Rendah:** Memastikan model dapat memberikan prediksi secara real-time, yang esensial untuk aplikasi seperti kendaraan otonom.
3. **Konsumsi Daya:** Mengurangi konsumsi energi pada perangkat bertenaga baterai.
4. **Kompatibilitas Hardware:** Menyesuaikan model dengan spesifikasi dan batasan perangkat keras tertentu.

D. Teknik Optimisasi Model

1. **Pruning:** Menghilangkan parameter atau koneksi yang tidak penting untuk mengurangi ukuran dan kompleksitas model.
2. **Quantization:** Mengonversi bobot model dari presisi tinggi (32-bit) ke format presisi rendah (8-bit) untuk mengurangi memori dan meningkatkan kecepatan inferensi.
3. **Knowledge Distillation:** Mentransfer pengetahuan dari model besar (teacher) ke model yang lebih kecil (student) untuk menciptakan model yang lebih ringan dengan performa serupa.
4. **Low-Rank Approximation:** Menggunakan pendekatan matriks kecil untuk mendekati matriks besar, mengurangi konsumsi memori.
5. **Hardware-Specific Compression:** Optimisasi model untuk perangkat keras tertentu seperti NVIDIA GPU atau Intel Hardware.

E. Trade-Off antara Akurasi, Performa, dan Penggunaan Sumber Daya

Optimisasi model melibatkan keseimbangan antara:

- **Akurasi:** Model dengan akurasi tinggi cenderung lebih kompleks dan memerlukan lebih banyak sumber daya.
- **Performa (Latency):** Kecepatan inferensi penting untuk aplikasi real-time, namun sering mengorbankan akurasi.
- **Penggunaan Sumber Daya:** Efisiensi penggunaan CPU, memori, dan penyimpanan penting untuk perangkat dengan keterbatasan sumber daya.

III. Zero-Shot Learning (ZSL)

A. Definisi dan Klasifikasi

Zero-Shot Learning (ZSL) adalah metode di mana model AI dapat mengklasifikasikan kelas yang tidak pernah dilihat selama pelatihan. Ada dua jenis utama ZSL:

1. **Inductive ZSL:** Model dilatih hanya pada kelas yang terlihat tanpa akses ke data dari kelas yang tidak terlihat.
2. **Transductive ZSL:** Model memiliki akses ke atribut atau contoh data tak berlabel dari kelas yang tidak terlihat selama pelatihan.

Generalized Zero-Shot Learning (GZSL) memperluas ZSL dengan menguji model pada kombinasi kelas yang terlihat dan tidak terlihat, mencerminkan skenario dunia nyata yang lebih kompleks.

B. Penerapan dalam Computer Vision

ZSL memerlukan informasi tambahan selain fitur visual, seperti semantic embeddings, untuk mengklasifikasikan objek dari kelas yang tidak terlihat. Pendekatan utama dalam ZSL:

1. **Embedding-Based Methods:** Proyeksikan fitur visual dan semantik ke ruang embedding bersama untuk melakukan klasifikasi berdasarkan kesamaan.
2. **Generative-Based Methods:** Menggunakan model generatif untuk membuat data sintetis dari kelas yang tidak terlihat, memungkinkan pelatihan model pada data sintetis tersebut.

C. Perbandingan dengan Metode Lain

1. **Transfer Learning:** ZSL adalah subset dari transfer learning yang berfokus pada transfer pengetahuan dari kelas terlihat ke kelas tidak terlihat.
2. **Domain Adaptation (DA):** Berbeda dengan ZSL yang tidak menggunakan data dari kelas tidak terlihat, DA memerlukan data label dari domain target.
3. **Open Set Recognition (OSR):** OSR hanya mengidentifikasi apakah data baru berasal dari kelas yang terlihat atau tidak, tanpa menentukan label.
4. **Out-of-Distribution (OOD) Detection:** Fokus pada mendeteksi data yang menyimpang dari distribusi pelatihan.
5. **Open Vocabulary Learning (OVL):** Ekstensi dari ZSL yang mencakup kemampuan untuk belajar dari data terbatas untuk kelas yang tidak terlihat dan terlihat serta menangani tugas baru yang tak terbatas.

IV. Etika dan Bias dalam Computer Vision

A. Studi Kasus: ImageNet Roulette

ImageNet Roulette adalah eksperimen yang menyoroti bias dalam dataset ImageNet. Aplikasi ini mengklasifikasikan gambar orang menggunakan subset "person" dari ImageNet, sering kali menghasilkan label yang ofensif dan stereotipikal.

Masalah Utama:

1. **Label Bias:** Label seperti "kriminal," "pecandu," dan lainnya muncul karena struktur WordNet yang digunakan oleh ImageNet.
2. **Bias Data Pelatihan:** Dataset mencerminkan bias sosial dan stereotip, menyebabkan model mengklasifikasikan secara merugikan individu atau kelompok tertentu.
3. **Kurangnya Pengawasan Etis:** Data yang diambil tanpa kontrol ketat menyebabkan inklusi label yang tidak pantas.

B. Implikasi dan Dampak

1. **Dampak pada Individu dan Kelompok:**
 - **Profil Rasial dan Gender:** Memperkuat stereotip dan diskriminasi.
 - **Pelanggaran Privasi:** Mengaburkan wajah dan menggunakan teknik pengaburan untuk melindungi identitas.
2. **Dampak pada Komunitas AI:**
 - **Transparansi dan Etika:** Mendorong perlunya transparansi dan etika dalam pembuatan dan penggunaan dataset.
 - **Responsibilitas Komunitas:** Menyoroti tanggung jawab komunitas AI untuk memastikan dataset yang adil dan tidak bias.

C. Solusi dan Mitigasi oleh Tim ImageNet

1. **Penghapusan Label Sensitif:** Menghapus 1.593 synsets yang ofensif dan sensitif dari dataset.
2. **Penghapusan Konsep Non-Imageable:** Menghapus synsets yang tidak dapat diklasifikasikan secara visual.
3. **Diversitas dan Representasi:** Menyeimbangkan representasi gender, ras, dan usia melalui analisis demografis.

4. **Kepatuhan Privasi:** Mengaburkan wajah dalam dataset untuk melindungi identitas individu tanpa mengorbankan akurasi model.

V. Upaya Hugging Face dalam Etika dan Masyarakat

A. Prinsip Democratizing Good Machine Learning

Hugging Face berkomitmen untuk mendemokratisasi *Good ML* melalui prinsip kolaborasi, transparansi, dan tanggung jawab.

1. Kolaborasi:

- **Model Cards:** Dokumentasi transparan tentang model, termasuk penggunaan, batasan, dan metrik performa.
- **Evaluasi:** Memungkinkan evaluasi model pada dataset terbuka tanpa perlu menulis kode.
- **Diskusi Komunitas:** Forum untuk feedback dan kontribusi.
- **Grup Discord:** Diskusi lintas domain seperti NLP, reinforcement learning, game development, audio, dan computer vision.

2. Transparansi:

- **Ethical Charter:** Charter etika untuk proyek multimodal.
- **AI Policy:** Respons terhadap laporan kebijakan AI nasional.

3. Tanggung Jawab:

- **Proyek Pendidikan:** Lokakarya, meetups, dan konferensi untuk mendidik berbagai kalangan.
- **Data Measurement Tool:** Alat inter aktif untuk mengevaluasi metrik dataset secara otomatis.

B. Kategori Hugging Face Spaces Berdasarkan Aspek Etika ML

Hugging Face Spaces dikategorikan ke dalam enam aspek etika utama untuk memastikan aplikasi machine learning yang bertanggung jawab dan berkelanjutan:

1. 🛠️ Rigorous

- **Fokus:** Memastikan privasi, transparansi, dan mitigasi bias.
- **Contoh Proyek:**
 - **Diffusion Bias Explorer:** Membandingkan bias dalam model teks-ke-gambar seperti SD 1.4, SD 2.0, dan DALL-E 2.
- **Teknik:**
 - Penghapusan overfitting dan memorisasi data pelatihan.
 - Detoksifikasi model bahasa.

2. 🤝 Consentful

- **Fokus:** Menghormati otonomi pengguna dan privasi.

- **Contoh Proyek:**
 - **Does CLIP Know My Face:** Menguji apakah model CLIP dapat memprediksi nama pengguna berdasarkan gambar yang diunggah.
 - **Photoguard:** Melindungi gambar dari manipulasi oleh model editing foto berbasis ML.
- 3. 🧐 **Socially Conscious**
 - **Fokus:** Menggunakan ML untuk kebaikan sosial.
 - **Contoh Proyek:**
 - **Socratic Models Image Captioning:** Meningkatkan aksesibilitas melalui captioning otomatis.
 - **Model untuk perubahan iklim dan revitalisasi bahasa Indigen.**
- 4. 🌍 **Sustainable**
 - **Fokus:** Menciptakan ML yang ramah lingkungan.
 - **Contoh Proyek:**
 - **EfficientFormer:** Model yang hemat energi.
 - **EfficientNetV2:** Model yang mengurangi jejak karbon tanpa mengorbankan kualitas.
- 5. 👥 **Inclusive**
 - **Fokus:** Memperluas partisipasi dan representasi dalam AI.
 - **Contoh Proyek:**
 - **Promptist Demo:** Solusi no-code untuk interaksi AI.
 - **Dukungan untuk bahasa dan dataset yang kurang terwakili.**
- 6. 🤔 **Inquisitive**
 - **Fokus:** Mengkritisi dan menantang norma dalam AI.
 - **Contoh Proyek:**
 - **PAIR: Datasets Have Worldviews:** Mengeksplorasi bias dalam pembuatan dataset.
 - **Proyek yang merefleksikan perspektif masyarakat adat dan LGBTQIA2S+ dalam AI.**

VI. Kesimpulan dan Rekomendasi

A. Kesimpulan

Pengembangan model AI yang efisien, etis, dan inklusif memerlukan pendekatan multidimensional yang mencakup optimisasi teknis, pengelolaan bias, dan penerapan prinsip-prinsip etika. Studi kasus seperti ImageNet Roulette menunjukkan dampak serius dari bias dalam dataset, sementara inisiatif komunitas seperti Hugging Face menawarkan solusi inovatif untuk menciptakan AI yang lebih adil dan bertanggung jawab. Dengan menerapkan teknik optimisasi yang tepat dan memastikan dataset yang bebas bias, kita dapat mengembangkan model AI yang tidak hanya akurat tetapi juga etis dan inklusif.

B. Rekomendasi

1. Optimisasi Model:

- Terapkan teknik seperti pruning, quantization, dan knowledge distillation untuk meningkatkan efisiensi model pada perangkat terbatas.
- Lakukan evaluasi menyeluruh untuk memastikan keseimbangan antara akurasi dan performa.

2. Mitigasi Bias:

- Audit dataset secara rutin untuk mengidentifikasi dan menghapus label yang bias.
- Lakukan analisis demografis untuk memastikan representasi yang adil dalam dataset.

3. Etika dalam AI:

- Libatkan komunitas dalam pengembangan AI untuk meningkatkan transparansi dan akuntabilitas.
- Edukasi tim pengembangan tentang pentingnya etika dan dampak sosial dari teknologi AI.

4. Kolaborasi dengan Komunitas:

- Bangun ruang diskusi terbuka seperti yang dilakukan oleh Hugging Face.
- Libatkan pemangku kepentingan dari berbagai latar belakang untuk memperkaya perspektif dalam pengembangan teknologi.

5. Inisiatif Berkelanjutan dan Inklusif:

- Fokus pada pembuatan ML yang ramah lingkungan dengan mengurangi jejak karbon.
- Tingkatkan representasi kelompok yang kurang terlayani melalui pembuatan dataset yang beragam dan mendukung bahasa yang kurang terwakili.

VII. Referensi

1. Goodfellow, I., et al. (2014). Generative Adversarial Networks. *NIPS*.
2. Xian, Y., et al. (2018). Zero-shot Learning - A Comprehensive Evaluation of the Good, the Bad and the Ugly. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
3. Mikolov, T., et al. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.
4. Pennington, J., et al. (2014). GloVe: Global Vectors for Word Representation. *EMNLP*.
5. Radford, A., et al. (2021). Learning Transferable Visual Models From Natural Language Supervision. *ICML*.
6. Palatucci, M., et al. (2009). Zero-Shot Learning Through Cross-Modal Transfer. *NeurIPS*.
7. Lampert, C.H., et al. (2009). Attribute-Based Classification for Zero-Shot Visual Object Categorization. *CVPR*.
8. Russakovsky, O., et al. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*.