



SISTEM REKOMENDASI FASHION OUTFIT MENGGUNAKAN DEEP NEURAL NETWORK

Marceleo Pratama Putra, Hasrullah
STMIK Borneo Internasional Balikpapan

Email: marceleo.pputra@gmail.com, hasrullah@permataabadi.co.id

Abstrak Industri fashion digital mengalami pertumbuhan pesat dengan meningkatnya kebutuhan sistem rekomendasi outfit yang personal dan akurat. Penelitian ini mengkaji berbagai pendekatan Deep Neural Network (DNN) untuk sistem rekomendasi outfit dengan fokus pada arsitektur Convolutional Neural Network (CNN), teknik pembelajaran metrik, dan personalisasi pengguna. Studi literatur mencakup sepuluh penelitian terkini yang mengimplementasikan berbagai arsitektur seperti VGGNet, ResNet-50, AlexNet, dan BN-Inception untuk ekstraksi fitur visual dari gambar fashion. Metode yang dikaji meliputi fine-tuning untuk personalisasi, Matrix Factorization untuk menangani masalah cold-start, ukuran kemiripan untuk rekomendasi berbasis konten, serta integrasi dengan atribut fisik pengguna seperti warna kulit dan bentuk tubuh. Hasil kajian menunjukkan bahwa pendekatan hybrid yang menggabungkan fitur gambar laten dengan fitur log laten menghasilkan akurasi terbaik hingga 92%, sementara model berbasis ResNet-50 dengan pengklasifikasi biner mencapai performa 84%. Integrasi pembelajaran kepribadian dengan atribut fisik menggunakan jaringan message-passing dan Transformer mencapai akurasi 93% pada dataset SOP. Penelitian ini memberikan wawasan komprehensif tentang perkembangan terkini dalam sistem rekomendasi outfit dan arah pengembangan masa depan yang meliputi augmented reality untuk virtual try-on, pembelajaran multi-modal, dan sistem memori adaptif. Kata Kunci: <i>Sistem Rekomendasi Outfit, Deep Neural Network, Convolutional Neural Network, Pengenalan Fashion, Personalisasi Fashion</i> Abstract <i>The digital fashion industry has experienced rapid growth, accompanied by an increasing demand for personalized and accurate outfit recommendation systems. This study examines various Deep Neural Network (DNN) approaches for outfit recommendation, with a focus on Convolutional Neural Network (CNN) architectures, metric learning techniques, and user personalization. The literature review covers ten recent studies that implement architectures such as VGGNet, ResNet-50, AlexNet, and BN-Inception for visual feature extraction from fashion images. The reviewed methods include fine-tuning strategies for personalization, Matrix Factorization to address the cold-start problem, similarity measures for content-based recommendation, and the integration of user physical attributes such as skin tone and body shape. The results indicate that hybrid</i>	Article History Received: Desember 2025 Reviewed: Desember 2025 Published: Desember 2025 Plagiarism Checker No 799 Prefix DOI : Prefix DOI : 10.8734/Kohesi.v1i2.365 Copyright : Author Publish by : Kohesi  This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License
---	---



approaches combining latent image features with latent log features achieve the highest accuracy of up to 92%, while ResNet-50-based models with binary classifiers reach a performance of 84%. Furthermore, the integration of personality learning with physical attributes using message-passing networks and Transformer architectures achieves an accuracy of 93% on the SOP dataset. This study provides a comprehensive overview of recent advancements in outfit recommendation systems and highlights future research directions, including augmented reality for virtual try-on, multimodal learning, and adaptive memory-based systems.

Keyword: *Outfit Recommendation System, Deep Neural Network, Convolutional Neural Network, Fashion Recognition, Fashion Personalization*

1. PENDAHULUAN

Industri fashion global telah mengalami transformasi digital yang signifikan dalam dekade terakhir. Penjualan fashion online mencapai 370 miliar dolar di Amerika Serikat pada tahun 2017 dan 191 miliar euro di Eropa, dengan pertumbuhan tahunan mencapai 18% [4]. Di China, platform e-commerce seperti Taobao.com mencatat transaksi lebih dari 103 miliar RMB hanya dalam satu hari festival [4]. Pertumbuhan eksponensial ini menciptakan tantangan baru dalam membantu konsumen memilih outfit yang sesuai dari jutaan pilihan produk yang tersedia.

Fashion memiliki dampak yang sangat besar terhadap masyarakat modern [4]. Pakaian bukan hanya kebutuhan utilitarian, tetapi juga merupakan sarana ekspresi diri dan komunikasi personal yang kuat. Platform media sosial seperti Instagram, Facebook, dan Polyvore telah menjadi wadah bagi jutaan pengguna untuk berbagi foto fashion mereka dengan dunia [1]. Fenomena ini menciptakan kebutuhan mendesak akan sistem rekomendasi yang dapat memahami preferensi fashion individual dan memberikan saran outfit yang personal.

Sistem rekomendasi tradisional dalam domain fashion sebagian besar bergantung pada metode collaborative filtering yang menggunakan data historis rating pengguna [1,8]. Namun, pendekatan ini menghadapi beberapa keterbatasan fundamental. Pertama, masalah sparsity menjadi lebih serius dalam rekomendasi set outfit dibandingkan rekomendasi item individual, karena rating berhubungan dengan entri tensor user-item yang berdimensi tinggi [1]. Kedua, metode berbasis teks atau metadata manual memerlukan anotasi yang konsisten dan akurat, yang sulit dicapai dalam skala besar [3]. Ketiga, cold-start problem menjadi kritis dalam industri fashion di mana produk baru muncul setiap hari sesuai dengan tren dan musim [8].

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kajian literatur sistematis dengan fokus pada publikasi-publikasi terkini (2015-2024) yang membahas sistem rekomendasi outfit menggunakan deep neural networks. Sepuluh paper utama dipilih berdasarkan kriteria relevansi dengan topik penelitian, kualitas metodologi, kontribusi inovatif, dan hasil eksperimental yang tervalidasi.

2.1 Arsitektur Deep Neural Network untuk Rekomendasi Fashion

CNN merupakan komponen fundamental dalam hampir semua sistem rekomendasi outfit modern. Arsitektur CNN terdiri dari beberapa layer konvolusional yang mengekstraksi fitur hierarkis dari gambar input, diikuti oleh pooling layers untuk reduksi dimensionalitas, dan fully connected layers untuk klasifikasi atau embedding.

Arsitektur VGGNet

VGGNet, khususnya VGG-16, digunakan secara luas untuk klasifikasi fashion dan ekstraksi fitur. Khalid et al. mengimplementasikan VGG-16 untuk mengidentifikasi kategori garmen dengan mencapai akurasi 98% pada dataset fashion dengan 7 kategori (Shirt, T-Shirt, Skirt, Sweater, Heels, Pants, Shoes). Arsitektur VGG-16 terdiri dari 13 convolutional layers dan 3 fully connected layers, dengan filter size 3×3 dan max pooling 2×2 . Model ini menggunakan Local Response Normalization dan fungsi aktivasi ReLU.

Keunggulan VGG-16 terletak pada arsitekturnya yang sederhana namun dalam, dengan menggunakan multiple convolutional layers dengan filter kecil (3×3) yang dapat menangkap fitur kompleks secara efektif. Untuk rekomendasi fashion, VGG-16 dapat mengekstraksi fitur dari layer FC7 (4096 dimensi) yang kemudian digunakan untuk pencocokan kemiripan.

Arsitektur ResNet

Residual Networks (ResNet) memperkenalkan konsep skip connections yang memungkinkan pelatihan network yang sangat dalam tanpa mengalami masalah vanishing gradient. Huang & Huang menggunakan ResNet-50 sebagai feature extractor dalam sistem rekomendasi outfit mereka, menghasilkan feature vector 2048-dimensi dari layer pooling ke-5. ResNet-50 dipilih karena kemampuannya dalam menangkap fitur visual yang kompleks sambil mempertahankan efisiensi komputasi.

Rathod et al. juga mengimplementasikan ResNet-50 untuk rekomendasi fashion dengan virtual try-on, mencapai performa yang solid dalam mengidentifikasi tipe dan warna pakaian dari gambar. Arsitektur ResNet dengan identity mapping-nya memungkinkan aliran gradient yang lebih baik selama backpropagation, sehingga dapat melatih model yang lebih dalam dan akurat.

AlexNet dan BN-Inception

Tuinhof et al. melakukan studi komparatif antara AlexNet dan Batch Normalized Inception (BN-Inception) untuk rekomendasi produk fashion. Hasil pretraining pada dataset DeepFashion menunjukkan bahwa BN-Inception mencapai akurasi 63% dan top-5 akurasi 84% untuk klasifikasi kategori, melampaui AlexNet yang mencapai akurasi 57% dan top-5 akurasi 79%.

BN-Inception mengatasi masalah internal covariate shift melalui batch normalization, yang membuat pelatihan lebih stabil dan memungkinkan penggunaan learning rate yang lebih tinggi. Arsitektur Inception menggunakan multiple parallel convolutional operations dengan filter sizes yang berbeda, memungkinkan network untuk menangkap fitur pada multiple scales secara simultan.



2.2 Transfer Learning dan Fine-tuning

Transfer learning merupakan teknik krusial dalam rekomendasi fashion mengingat keterbatasan dataset yang tersedia. Model-model CNN umumnya diinisialisasi dengan bobot yang telah di-pretrain pada dataset ImageNet yang berisi 14 juta gambar dari 1000 kategori.

Strategi Pelatihan Dua Tahap

He & Hu mengusulkan strategi pelatihan dua tahap untuk personalisasi. Tahap pertama melatih model kompatibilitas umum menggunakan outfit dari semua user yang dicampur dengan neutral outfits (kombinasi acak dari items). Tahap kedua melakukan fine-tuning dengan data spesifik setiap user untuk mempelajari preferensi personal. Strategi ini menghasilkan peningkatan mean NDCG dari 0.59 (tahap satu) menjadi 0.81 (fine-tuning seluruh network tahap dua) pada dataset Polyvore.

Fine-tuning dapat dilakukan dengan dua strategi: fine-tune seluruh network termasuk feature extractor, atau freeze feature extractor dan hanya fine-tune matching network. Strategi pertama menghasilkan akurasi lebih tinggi namun memerlukan komputasi lebih besar, sementara strategi kedua lebih efisien untuk deployment.

2.3 Arsitektur Network Khusus

Arsitektur FashionNet

FashionNet yang diusulkan oleh He & Hu mengeksplorasi tiga alternatif arsitektur untuk pembelajaran kompatibilitas fashion:

1. **FashionNet A:** Menggabungkan gambar items menjadi single image dengan 9 color channels, kemudian memproses dengan VGGNet secara end-to-end.
2. **FashionNet B:** Memisahkan pembelajaran fitur dan pemodelan kompatibilitas. Items diproyeksikan ke common latent space menggunakan shared CNN, kemudian features di-concatenate dan diproses dengan fully connected layers untuk penilaian kompatibilitas.
3. **FashionNet C:** Mendekomposisi kompatibilitas high-order menjadi hubungan berpasangan. Model ini mengasumsikan kompatibilitas outfit ditentukan oleh kompatibilitas berpasangan antar items, sehingga lebih efisien dan mengatasi curse of dimensionality.

Hasil eksperimen menunjukkan FashionNet C mencapai performa terbaik dengan mean NDCG 0.81 dan 8.18 positive outfits dalam top-10 results, memvalidasi efektivitas pendekatan dekomposisi berpasangan.

P-Net (Personality Network)

Mo et al. mengusulkan P-Net yang mengintegrasikan pembelajaran kepribadian dengan kompatibilitas atribut fisik. Arsitektur P-Net terdiri dari:

1. **Feature Encoder:** Backbone ResNet-18 yang di-pretrain pada ImageNet untuk ekstraksi fitur visual



2. **Message Passing Network:** Memodelkan relasi global antar outfits dalam batch menggunakan mekanisme multi-head attention
3. **Style-aware Masking:** Mempelajari style-specific embeddings untuk perbandingan diskriminatif
4. **Transformer Module:** Memodelkan relasi antara feature embeddings dan physical attribute embeddings
5. **User Anchors:** Learnable embeddings yang merepresentasikan sifat kepribadian user

P-Net mencapai akurasi 93.09% dan NDCG 0.9434 pada dataset SOP, mendemonstrasikan keunggulan pendekatan yang mengintegrasikan berbagai aspek.

2.4 Metode Pembelajaran dan Optimasi

Pendekatan Metric Learning

Beberapa penelitian menggunakan arsitektur Siamese network untuk mempelajari metrik kemiripan. Siamese networks menggunakan weight-sharing antara dua atau lebih identical subnetworks, memproses input pairs secara parallel. Keuntungan Siamese networks adalah dapat langsung dioptimasi untuk tugas similarity, bukan klasifikasi.

Triplet loss digunakan untuk mempelajari embedding space di mana anchor-positive pairs memiliki distance lebih kecil dibanding anchor-negative pairs. He & Hu menggunakan ranking loss yang dioptimasi untuk pengurutan preferensial, dimana formulasi ini memastikan positive outfits mendapat score lebih tinggi dari negative outfits.

Ukuran Kemiripan

Cosine similarity banyak digunakan untuk mengukur kesamaan antar feature vectors, mengukur cosine dari sudut antara dua vectors dengan nilai berkisar dari -1 hingga 1. Nilai mendekati 1 menunjukkan similarity tinggi.

Euclidean distance mengukur straight-line distance antara dua points dalam feature space. Distance yang lebih kecil menunjukkan similarity yang lebih tinggi.

Algoritma k-Nearest Neighbors (k-NN) digunakan untuk menemukan item serupa dalam feature space. Tuinhof et al. menggunakan k-NN dengan implementasi ball-tree untuk pencarian nearest neighbor yang efisien pada dataset besar. Rathod et al. menggabungkan fitur CNN dengan k-NN untuk rekomendasi fashion, memanfaatkan CNN untuk ekstraksi fitur dan k-NN untuk pengambilan berbasis kemiripan.

Pendekatan Klasifikasi

Untuk klasifikasi biner, Huang & Huang menggunakan binary classifier untuk menilai apakah outfit adalah "baik" atau "buruk". Model menggunakan 2-layer Multi-Layer Perceptron (MLP) dengan input berupa concatenated feature vectors dari semua items dalam outfit. Output layer menggunakan sigmoid activation untuk binary classification, dengan threshold 0.5 untuk decision boundary.

Untuk klasifikasi multi-kelas, fungsi aktivasi softmax digunakan pada output layer. Bag et al. mengklasifikasikan warna kulit ke dalam 3 tipe (Type-I, Type-II, Type-III) menggunakan CNN dengan softmax classifier, mencapai akurasi 92%.



Untuk klasifikasi multi-label, prediksi kompatibilitas atribut fisik merupakan masalah multi-label classification. Mo et al. menggunakan Transformer encoder untuk memodelkan relasi antara feature embeddings dan label embeddings, dengan binary cross entropy loss untuk setiap label.

2.5 Personalisasi dan Adaptasi

Memori Jangka Panjang dan Jangka Pendek

Padilla Reyes et al. mengusulkan sistem rekomendasi adaptif dengan mekanisme memori ganda. Memori jangka pendek diperbarui setiap kali user berinteraksi dengan sistem menggunakan mekanisme ranking update, dimana setiap outfit memiliki quality score Q yang diupdate berdasarkan user selection. Memori jangka panjang berupa DNN yang dilatih ulang secara periodik setelah n interaksi untuk mempelajari pola preferensi jangka panjang.

Formula update untuk memori jangka pendek adalah $Q = Q + ((1-Q) \times \alpha)$ jika $\alpha > 0$ (reward), dan $Q = Q + (Q \times \alpha)$ jika $\alpha \leq 0$ (penalty), dimana α dihitung menggunakan fungsi sigmoid berdasarkan jarak L2 antara selected outfit dan candidate outfit.

Matrix Factorization untuk Masalah Cold-start

Ito et al. mengusulkan pendekatan hybrid yang menggabungkan CNN dengan Matrix Factorization (MF) untuk mengatasi masalah cold-start. Untuk item baru tanpa histori klik, sistem mengekstrak fitur gambar laten menggunakan CNN yang dilatih untuk klasifikasi warna, tipe, dan pola, menemukan item yang ada dengan fitur gambar laten paling mirip, dan melengkapi fitur log laten dari item baru dengan fitur log laten dari item serupa yang ada.

Pendekatan ini meningkatkan rasio item baru dari 3% (hanya DNN) menjadi 19% (DNN dengan pelengkap) sambil mempertahankan skor MAP@8 0.025. Matrix Factorization mendekomposisi matriks interaksi user-item $R(M \times N)$ menjadi matriks user $U(M \times K)$ dan matriks item $P(N \times K)$, dimana faktor laten item P_j kemudian digunakan sebagai fitur log laten yang merepresentasikan kemungkinan klik.

Integrasi Atribut Fisik

Bag et al. mengimplementasikan identifikasi warna kulit menggunakan CNN untuk rekomendasi outfit personal. Sistem mengidentifikasi warna kulit berdasarkan skala Fitzpatrick (6 tipe) dan merekomendasikan outfit yang kompatibel dengan warna kulit user. Model CNN mencapai akurasi 92% pada klasifikasi warna kulit.

Sistem juga mempertimbangkan bentuk tubuh (pria: Ectomorph, Mesomorph, Endomorph; wanita: Round/Apple, Pear, Inverted Triangle, Rectangular, Hourglass) untuk rekomendasi outfit. Integrasi dengan IBM Watson Assistant digunakan untuk mengumpulkan informasi user (jenis kelamin, kelompok usia, ukuran) sebelum menghasilkan rekomendasi.

Mo et al. mengintegrasikan 7 atribut fisik: bentuk tubuh, warna kulit, gaya rambut, warna rambut, tinggi badan, ukuran payudara, dan kontras warna dalam P-Net. Transformer encoder memodelkan relasi antara fitur visual dan atribut fisik untuk prediksi kompatibilitas. Model mencapai mAP 41.60% dan OF1 56.57% pada prediksi kompatibilitas fisik.



Rekomendasi Berbasis Kepribadian

P-Net mengintegrasikan sifat kepribadian MBTI (16 tipe) dengan preferensi gaya fashion. Sifat kepribadian diprediksi dari data media sosial menggunakan fitur psikolinguistik dan embeddings model bahasa. Pakar fashion memetakan 16 tipe MBTI ke gaya fashion berbeda, menciptakan pemetaan kepribadian-gaya.

User anchors (learnable embeddings) digunakan untuk merepresentasikan kepribadian user dalam embedding space. Message passing network dengan mekanisme multi-head attention mempelajari hubungan global antar outfits, dan strategi masking sadar gaya membuat embeddings diskriminatif untuk gaya berbeda.

2.6 Dataset dan Evaluasi

Dataset Fashion

Dataset Polyvore merupakan dataset paling populer untuk penelitian rekomendasi outfit, berisi outfit yang dibuat oleh users pada platform Polyvore. He & Hu menggunakan dataset dengan 800 users dan 21,623 positive outfits. Dataset menyediakan informasi seperti nama outfit, items, gaya serupa, komentar, dan jumlah like.

Dataset DeepFashion adalah dataset skala besar dengan lebih dari 800,000 gambar fashion dan anotasi lengkap termasuk kategori, landmarks, dan atribut. Tuinhof et al. menggunakan subset DeepFashion Attribute Prediction (289,222 gambar) untuk pretraining.

Dataset Fashion Product Images dari Kaggle dengan 44,000 produk, masing-masing dengan multiple category labels, deskripsi, dan gambar berkualitas tinggi. Dataset ini banyak digunakan untuk penelitian karena anotasi komprehensif dan kategori produk yang beragam.

Dataset SOP (Stylish Outfit of Personality) yang baru dikembangkan oleh Mo et al., berisi 969 users dengan 16 anotasi kepribadian MBTI dan pasangan pemetaan kepribadian-outfit. Total 161,014 pasangan positif/negatif pada training set dan 151,911 pasangan pada testing set. Dataset ini unik karena mengintegrasikan informasi psikologis dengan data fashion.

Metrik Evaluasi

Akurasi dan Loss adalah metrik dasar untuk tugas klasifikasi. Khalid et al. melaporkan akurasi 86% dan loss 0.51 untuk model mereka. Bag et al. mencapai akurasi 92% untuk klasifikasi warna kulit.

Mean NDCG (Normalized Discounted Cumulative Gain) adalah metrik untuk mengevaluasi kualitas ranking, dimana $yp(i)$ adalah skor relevansi dari item pada posisi i . He & Hu mencapai mean NDCG 0.81 dengan FashionNet C.

MAP (Mean Average Precision) mengukur precision across different recall levels. Ito et al. melaporkan skor MAP@8 0.025 untuk model hybrid CNN+MF.

AUC (Area Under ROC Curve) mengukur kemampuan diskriminatif model. Ito et al. mencapai AUC 0.71 dengan model yang diusulkan.

Top-K Accuracy mengukur apakah true positive ada dalam top-K rekomendasi. He & Hu melaporkan 8.18 positive outfits dalam top-10 results.

Metrik Kompatibilitas Fisik yang digunakan Mo et al. meliputi CP (Category-wise Precision), CR (Category-wise Recall), CF1 (Category-wise F1), OP (Overall Precision), OR (Overall Recall), dan OF1 (Overall F1) untuk mengevaluasi prediksi kompatibilitas atribut fisik.

3. HASIL DAN PEMBAHASAN

3.1 Perbandingan Performa Berbagai Arsitektur

Tabel berikut merangkum performa berbagai arsitektur CNN dalam tugas rekomendasi fashion:

Arsitektur	Dataset	Akurasi	NDCG / MAP	Refrensi
VGG-16	Fashion-7K	98%	-	[3]
ResNet-50	Polyvore	84%	-	[4]
AlexNet	Fashion	79% (top-5)	-	[5]
BN-Inception	Fashion	84% (top-5)	-	[5]
ResNet-18	SOP	84.32%	0.843	[10]

VGG-16 menunjukkan performa sangat baik untuk tugas klasifikasi dengan kategori dataset yang terbatas (7 kategori), mencapai akurasi 98%. Namun, untuk tugas yang lebih kompleks seperti pembelajaran kompatibilitas outfit, arsitektur yang lebih dalam seperti ResNet menunjukkan keunggulan.

ResNet-50 memberikan keseimbangan baik antara kedalaman model dan efisiensi komputasi, mencapai akurasi 84% pada klasifikasi biner outfit. Skip connections dalam arsitektur ResNet memungkinkan pelatihan efektif dari network yang sangat dalam tanpa masalah degradasi.

BN-Inception melampaui AlexNet secara konsisten, dengan peningkatan sebesar 5% pada akurasi top-5. Batch normalization dalam BN-Inception memungkinkan konvergensi lebih cepat dan generalisasi lebih baik. Modul Inception dengan konvolusi multi-skala paralel juga lebih efektif dalam menangkap pola visual beragam dalam gambar fashion.

Arsitektur Hybrid

Ito et al. mendemonstrasikan keunggulan pendekatan hybrid. Perbandingan performa:

Model	MAP@8	AUC	Rasio Item Baru
Hanya MF	0.022	0.66	0%
Hanya Gambar	0.017	0.71	16%
DNN (CNN+MF)	0.025	0.71	3%
DNN + Pelengkap	0.025	0.71	19%

Model hybrid mencapai peningkatan 14% dibanding hanya MF dan peningkatan 56% dibanding model hanya gambar dalam MAP@8. Yang krusial, strategi pelengkap

meningkatkan rasio item baru dari 3% menjadi 19% tanpa mengorbankan akurasi, secara efektif mengatasi masalah cold-start.

Varian Fashion Net

He & Hu membandingkan tiga arsitektur FashionNet:

Model	Mean NDCG	Top-10 Outfit Positif
FashionNet A	0.707	6.60
FashionNet B	0.796	7.83
FashionNet C	0.812	8.18

FashionNet C melampaui varian lainnya dengan margin 5.5% dalam mean NDCG dan 23.9% dalam jumlah outfit positif top-10. Keunggulan FashionNet C berasal dari strategi dekomposisi berpasangan yang mengurangi curse of dimensionality, memudahkan pembelajaran dari data yang terbatas, dan lebih intuitif dalam memodelkan aturan kompatibilitas fashion.

3.2 Efektivitas Personalisasi

He & Hu mengevaluasi berbagai strategi fine-tuning:

Strategi	Mean NDCG	Top-10 Positif
Awal (tanpa pelatihan)	0.322	1.22
Tahap satu (umum)	0.665	5.61
Tahap dua (parsial)	0.784	7.61
Tahap dua (menyeluruh)	0.812	8.12
Tahap dua (langsung)	0.615	5.30

Temuan kunci menunjukkan bahwa pelatihan dua tahap secara dramatis meningkatkan performa dibanding pelatihan dari awal (fine-tuning langsung), dengan peningkatan 19.7% dalam NDCG. Pelatihan umum (tahap satu) memberikan baseline kuat yang penting untuk personalisasi efektif. Fine-tuning seluruh network (menyeluruh) menghasilkan performa terbaik, namun fine-tuning parsial (freeze feature network) lebih efisien untuk deployment dengan sedikit pengorbanan akurasi.

Sistem Memori Adaptif

Padilla Reyes et al. mendemonstrasikan efektivitas pendekatan memori ganda. Pembaruan memori jangka pendek menunjukkan bahwa sistem mengupdate skor Q secara real-time berdasarkan interaksi pengguna. Hasil eksperimental menunjukkan bahwa setelah 20 interaksi memilih outfit dari kategori "Terburuk", skor Q dari outfit yang dipilih meningkat signifikan, outfit serupa juga mendapat reward (berbasis kedekatan), dan sistem berhasil mengadaptasi papan rekomendasi dari pelabelan profesional ke preferensi pengguna.

Pembelajaran memori jangka panjang dimana DNN dilatih ulang secara periodik menangkap pola preferensi jangka panjang. Visualisasi menunjukkan bahwa outfit yang awalnya dalam kategori "Terburuk" berpindah ke kategori "Terbaik" setelah pemilihan pengguna yang konsisten, mendemonstrasikan pembelajaran preferensi yang berhasil.

Personalisasi Berbasis Kepribadian

Mo et al. mencapai terobosan dalam rekomendasi sadar kepribadian:

Model	ACC	NDCG	mAP	CF1	OF1
Baseline Resnet18	84.32%	0.843	31.40	15.34	46.81
P-Net (tanpa msp)	63.43%	0.711	41.13	35.21	58.85
P-Net (tanpa smk)	80.75%	0.841	42.38	38.81	55.69
P-Net (lengkap)	93.09%	0.943	41.60	35.26	56.57

Message passing network (msp) krusial untuk menangkap hubungan outfit global, berkontribusi peningkatan akurasi 29.66%. Style-aware masking (smk) meningkatkan daya diskriminatif, menambahkan peningkatan akurasi 12.34%. Pendekatan gabungan mencapai akurasi 93.09%, mendemonstrasikan sinergi antar komponen.

3.3 Integrasi Atribut Fisik

Bag et al. mengimplementasikan sistem identifikasi warna kulit dengan metrik performa sebagai berikut: Akurasi 92%, Sensitivitas 97.7%, dan Spesifisitas 99%. Perkembangan pelatihan menunjukkan peningkatan akurasi setiap epoch dari 83.1% (Epoch 2), 83.8% (Epoch 4), 84.2% (Epoch 6), 84.8% (Epoch 8), hingga stabil di 84.4% (Epoch 10) dengan akurasi akhir 92%. Sistem berhasil mengidentifikasi warna kulit dan merekomendasikan warna outfit yang kompatibel berdasarkan klasifikasi skala Fitzpatrick.

Mo et al. mengevaluasi kompatibilitas fisik terhadap 7 atribut:

Model	mAP	CP	CR	CF1	OP	OR	OF1
Resnet18	31.40	12.44	20.00	15.34	59.01	38.79	46.81
CSN	37.76	39.04	30.95	34.53	57.86	48.82	52.96
P-Net	41.60	39.11	32.10	35.26	63.71	50.86	56.57

P-Net mencapai performa terbaik secara keseluruhan dengan skor mAP 41.60 dan OF1 56.57. Transformer encoder secara efektif memodelkan hubungan antara fitur visual dan atribut fisik, memungkinkan prediksi kompatibilitas yang akurat.

3.4 Solusi Masalah Cold-Start

Ito et al. mengatasi masalah cold-start dengan pelengkapan fitur laten. Kemampuan rekomendasi item baru menunjukkan bahwa MF Tradisional menghasilkan 0% item baru direkomendasikan, Hanya gambar 16% item baru, DNN tanpa pelengkap 3% item baru, dan DNN dengan pelengkap 19% item baru.

Mekanisme yang digunakan adalah pertama mengekstrak fitur gambar laten dari item baru menggunakan CNN, kedua menemukan item yang ada dengan fitur gambar paling



mirip (jarak kosinus minimum), ketiga menetapkan fitur log laten dari item mirip ke item baru, dan keempat menggunakan fitur gabungan untuk rekomendasi. Strategi ini berhasil memungkinkan rekomendasi item baru sambil mempertahankan akurasi dengan MAP@8 sebesar 0.025.

Tuinhof et al. mendemonstrasikan efektivitas transfer learning untuk mengatasi keterbatasan data. Hasil pretraining pada DeepFashion menunjukkan AlexNet mencapai akurasi kategori 57% dan akurasi tekstur 28%, sementara BN-Inception mencapai akurasi kategori 63% dan akurasi tekstur 32%. Hasil fine-tuning pada dataset Fashion menunjukkan BN-Inception mencapai akurasi kategori 87% dan akurasi tekstur 80%.

Pretraining pada dataset besar (DeepFashion) kemudian fine-tuning pada dataset target (Fashion) menghasilkan peningkatan performa yang signifikan. Transfer learning memungkinkan pembelajaran efektif bahkan dengan dataset target yang relatif kecil.

3.5 Efisiensi Komputasi

Untuk performa real-time, VGG-16 tidak melaporkan waktu pemrosesan, ResNet-50 memproses dalam sepersekian detik per gambar, dan P-Net memiliki kemampuan rekomendasi real-time. Rathod et al. mengimplementasikan sistem virtual try-on real-time yang dapat memproses frame video dan melapisi item pakaian dengan latensi minimal, mendemonstrasikan kelayakan deployment praktis.

Untuk efisiensi pelatihan, He & Hu melaporkan detail pelatihan untuk FashionNet C dimana Tahap 1 (model umum) memerlukan 18 epoch dan Tahap 2 (personalisasi) memerlukan 5-7 epoch per pengguna, menghasilkan penghematan waktu signifikan dibanding pelatihan dari awal. Padilla Reyes et al. mengimplementasikan mekanisme pembaruan efisien dengan pembaruan jangka pendek berupa pembaruan skor Q instan (tanpa pelatihan ulang) dan jangka panjang berupa pelatihan ulang DNN periodik (setiap n interaksi), menciptakan keseimbangan antara adaptabilitas dan biaya komputasi.

3.6 Analisis Clustering dan Embedding

Bhure et al. menggunakan clustering berbasis embedding untuk mengorganisir item fashion. Item fashion dikelompokkan berdasarkan kemiripan dalam ruang embedding. Sistem berhasil mengelompokkan item serupa, memungkinkan pengambilan produk serupa yang efisien, organisasi katalog produk besar yang lebih baik, dan peningkatan keragaman rekomendasi.

Performa pelatihan menunjukkan akurasi pelatihan pada Epoch 10 sebesar 84.4%, akurasi validasi sedikit lebih rendah menunjukkan generalisasi baik, dan penurunan loss konsisten dari 0.115 (Epoch 2) ke 0.065 (Epoch 10).



Mo et al. mengimplementasikan strategi masking sadar gaya dengan observasi kunci bahwa outfit dari gaya yang sama terletak berdekatan dalam ruang gaya yang sesuai, outfit dari gaya berbeda terpisah dalam embedding spesifik gaya, dan strategi memungkinkan perbandingan diskriminatif di berbagai dimensi gaya secara bersamaan. Regularisasi L2 pada mask gaya meningkatkan ketahanan dimana S adalah jumlah gaya dan ms adalah mask gaya ke- s .

3.7 Integrasi Multi-modal

Ito et al. berhasil mengintegrasikan fitur visual dengan data log klik. Dimensi fitur terdiri dari fitur gambar laten 3×1024 dimensi (dari 3 model CNN untuk warna, tipe, pola), fitur log laten $K=16$ dimensi (dari Matrix Factorization), dengan input gabungan total 3088 dimensi. Keuntungan arsitektur adalah fitur visual menangkap kemiripan penampilan, fitur log menangkap pola perilaku pengguna, fitur gabungan memungkinkan penilaian kemiripan komprehensif, dan operasi inner product secara efisien menggabungkan kedua jenis fitur.

Mo et al. mengintegrasikan berbagai modalitas data dengan sumber informasi berupa visual (gambar outfit diproses oleh ResNet-18), tekstual (sifat kepribadian dari analisis teks media sosial), fisik (7 atribut fisik seperti bentuk tubuh, warna kulit, dll.), dan perilaku (data interaksi pengguna-outfit).

Mekanisme integrasi menggunakan jaringan message passing untuk menangkap hubungan outfit, user anchors untuk merepresentasikan sifat kepribadian, transformer encoder untuk memodelkan hubungan visual-fisik, dan pelatihan end-to-end untuk mengoptimalkan semua komponen secara bersama.

3.8 Keterbatasan dan Tantangan

Untuk masalah ketidakseimbangan data, Mo et al. mengidentifikasi distribusi yang tidak seimbang dalam atribut fisik. Performa klasifikasi multi-label (mAP sekitar 41%) lebih rendah dari performa pembelajaran metrik (NDCG sekitar 94%), menunjukkan kesulitan dalam pembelajaran dari label yang tidak seimbang, kebutuhan strategi sampling yang lebih baik atau pembobotan ulang loss, dan manfaat potensial dari augmentasi data sintetis. Dataset SOP mungkin memiliki distribusi tipe kepribadian yang tidak seimbang, mempengaruhi kemampuan model untuk menggeneralisasi pada semua 16 tipe MBTI secara merata.

Untuk biaya komputasi, jaringan dalam memerlukan sumber daya GPU yang signifikan, personalisasi untuk banyak pengguna meningkatkan kebutuhan penyimpanan, dan adaptasi real-time memerlukan mekanisme pembaruan yang efisien. Ito et al. menggunakan implementasi ball-tree untuk pencarian k-NN yang efisien, mengurangi kebutuhan memori untuk deployment skala besar.



Untuk keterbatasan dataset, Mo et al. mencatat bahwa dataset SOP hanya dapat dibuka sebagian untuk melindungi privasi pengguna. Informasi kepribadian sensitif, membatasi ketersediaan dataset untuk komunitas riset. Sebagian besar dataset berfokus pada domain fashion tertentu (fashion Barat, demografi tertentu), berpotensi membatasi generalisasi ke konteks fashion budaya yang beragam.

4. KESIMPULAN

Berdasarkan kajian literatur terhadap sepuluh penelitian terkini dalam sistem rekomendasi outfit menggunakan deep neural networks, dapat disimpulkan beberapa temuan utama.

Berbagai arsitektur CNN telah terbukti efektif untuk rekomendasi fashion, dengan performa bervariasi tergantung kompleksitas tugas. VGG-16 sangat baik untuk tugas klasifikasi dengan kategori terbatas mencapai akurasi 98%, ResNet-50 menyediakan keseimbangan optimal antara kedalaman dan efisiensi mencapai akurasi 84-92%, BN-Inception superior untuk tugas yang memerlukan ekstraksi fitur multi-skala mencapai akurasi 87%, dan arsitektur yang lebih dalam umumnya berkinerja lebih baik untuk tugas pembelajaran kompatibilitas yang kompleks.

Personalisasi secara dramatis meningkatkan kualitas rekomendasi. Strategi pelatihan dua tahap meningkatkan NDCG dari 0.665 menjadi 0.812 (peningkatan 22%), fine-tuning spesifik pengguna penting untuk menangkap preferensi individual, sistem memori ganda (jangka pendek dan jangka panjang) memungkinkan adaptasi real-time yang efektif, dan pendekatan berbasis kepribadian mencapai akurasi tertinggi (93.09%) dengan menggabungkan sifat psikologis.

Beberapa strategi efektif untuk mengatasi masalah cold-start adalah pelengkap Matrix Factorization meningkatkan rasio item baru dari 3% ke 19%, transfer learning dari dataset besar memungkinkan pembelajaran efektif dengan data terbatas, kemiripan fitur visual dapat menggantikan data perilaku yang hilang, dan pendekatan hybrid yang menggabungkan berbagai sumber informasi paling efektif.

Integrasi berbagai modalitas data secara signifikan meningkatkan kualitas rekomendasi. Visual dan perilaku menghasilkan peningkatan 14% dibanding modalitas tunggal, visual dan atribut fisik memungkinkan penilaian kompatibilitas di luar penampilan, visual, kepribadian, dan fisik mencapai performa terbaik dengan akurasi 93%, dan optimasi end-to-end dari fitur multi-modal lebih superior daripada optimasi terpisah.

Pertimbangan karakteristik fisik pengguna meningkatkan utilitas praktis. Identifikasi warna kulit memungkinkan rekomendasi kompatibilitas warna dengan akurasi 92%, pertimbangan bentuk tubuh meningkatkan rekomendasi terkait kesesuaian, prediksi kompatibilitas multi-atribut (7 atribut) mencapai skor OF1 56.57%, namun prediksi kompatibilitas fisik masih menantang dengan mAP sekitar 41%.

Kajian ini memberikan beberapa kontribusi penting yaitu tinjauan komprehensif yang menyajikan analisis lengkap tentang perkembangan terkini dalam sistem rekomendasi fashion, mencakup pilihan arsitektur, strategi pelatihan, metode personalisasi, dan pendekatan evaluasi. Analisis komparatif memberikan perbandingan kuantitatif dan kualitatif dari berbagai pendekatan, membantu peneliti dan praktisi memilih metode



yang sesuai untuk kasus penggunaan spesifik. Identifikasi praktik terbaik dalam ekstraksi fitur menggunakan deep CNN dengan pemilihan arsitektur yang sesuai, personalisasi melalui pelatihan dua tahap dengan fine-tuning, penanganan cold-start dengan pendekatan hybrid visual-perilaku, dan integrasi multi-modal melalui kerangka kerja yang dapat dipelajari end-to-end. Analisis kesenjangan mengidentifikasi keterbatasan yang ada dan kesenjangan penelitian yang perlu diatasi dalam pekerjaan masa depan.

Beberapa keterbatasan dalam kajian ini adalah keragaman dataset dimana mayoritas penelitian menggunakan dataset dari konteks fashion Barat, berpotensi membatasi generalisasi ke preferensi fashion budaya yang beragam. Standardisasi evaluasi dimana studi berbeda menggunakan metrik evaluasi dan dataset yang berbeda, membuat perbandingan langsung menjadi menantang. Deployment dunia nyata dengan informasi terbatas tentang pengalaman deployment nyata, studi pengguna, dan metrik performa jangka panjang. Persyaratan komputasi dengan diskusi yang tidak memadai tentang biaya komputasi, efisiensi energi, dan pertimbangan skalabilitas untuk deployment produksi.

Berdasarkan analisis keterbatasan dan tren yang muncul, beberapa arah penelitian masa depan yang menjanjikan adalah integrasi augmented reality dengan virtual try-on dengan rendering foto-realistis, simulasi pakaian 3D real-time, ruang fitting virtual kolaboratif multi-pengguna, dan integrasi dengan kacamata AR untuk pengalaman di toko.

Pembelajaran mendalam multi-modal dengan model terpadu yang memproses data visual, tekstual, dan audio secara bersamaan, pengambilan cross-modal untuk tipe query yang fleksibel, generasi fashion berbasis panduan bahasa, dan rekomendasi outfit berbasis video.

Personalisasi lanjutan dengan rekomendasi kontekstual mempertimbangkan acara, cuaca, lokasi, pemodelan preferensi temporal untuk menangkap evolusi tren fashion, pemodelan pengaruh sosial untuk rekomendasi berbasis teman sebaya, dan rekomendasi sadar kondisi emosional.

Fashion berkelanjutan dengan sistem rekomendasi yang mempromosikan pilihan fashion berkelanjutan, optimasi capsule wardrobe, maksimalisasi mix-and-match untuk lemari pakaian yang ada, dan integrasi fashion bekas.

Keadilan dan etika dengan mengatasi bias dalam sistem rekomendasi, representasi fashion inklusif di berbagai demografi yang beragam, teknik personalisasi yang menjaga privasi, dan rekomendasi yang transparan dan dapat dijelaskan.

Peningkatan teknis dengan arsitektur efisien untuk deployment mobile, federated learning untuk pelestarian privasi, pembelajaran self-supervised untuk mengurangi kebutuhan anotasi, dan pembelajaran few-shot untuk adaptasi cepat ke tren fashion baru.

Pemodelan fisik yang ditingkatkan dengan integrasi pemindaian tubuh 3D, prediksi kesesuaian yang presisi menggunakan data antropometrik, pertimbangan properti material (elastisitas, jatuh, tekstur), dan model prediksi kenyamanan.



Aplikasi lintas domain dengan rekomendasi fashion untuk konteks budaya berbeda, adaptasi untuk domain khusus (pakaian olahraga, formal, kerja), rekomendasi fashion yang sesuai usia, dan fashion adaptif untuk orang dengan disabilitas.

Untuk praktisi dan pemain industri, platform e-commerce dapat mengimplementasikan pendekatan hybrid yang menggabungkan sinyal visual dan perilaku, berinvestasi dalam infrastruktur personalisasi untuk meningkatkan keterlibatan pengguna, mempertimbangkan kompatibilitas atribut fisik untuk mengurangi tingkat pengembalian, dan menerapkan fitur virtual try-on untuk meningkatkan pengalaman belanja online.

Pengecer fashion dapat memanfaatkan deep learning untuk manajemen inventori dan prediksi tren, mengimplementasikan sistem rekomendasi untuk pengalaman online dan di toko, menggunakan data pelanggan secara bertanggung jawab dengan perlindungan privasi yang sesuai, dan fokus pada rekomendasi fashion berkelanjutan.

Peneliti AI dapat memprioritaskan interpretabilitas dan penjelasan dalam sistem AI fashion, mengatasi kekhawatiran keadilan dan bias dalam pengembangan model, berkolaborasi dengan ahli fashion untuk integrasi pengetahuan domain, dan melakukan studi pengguna menyeluruh untuk memvalidasi peningkatan teknis.

Pembuat kebijakan dapat mengembangkan pedoman untuk AI etis dalam industri fashion, memastikan perlindungan privasi dalam sistem personalisasi, mempromosikan fashion berkelanjutan melalui insentif kebijakan, dan mengatasi potensi perpindahan pekerjaan dari otomasi AI.

Sistem rekomendasi outfit menggunakan deep neural networks telah menunjukkan kemajuan signifikan dalam beberapa tahun terakhir. Dari pendekatan berbasis konten sederhana hingga sistem multi-modal canggih yang mengintegrasikan sifat kepribadian dan atribut fisik, bidang ini terus berkembang pesat.

Faktor kesuksesan utama meliputi pemilihan arsitektur CNN yang sesuai untuk ekstraksi fitur, strategi personalisasi efektif untuk menangkap preferensi individual, pendekatan hybrid untuk mengatasi cold-start dan sparsitas data, integrasi multi-modal untuk pemahaman holistik, dan pertimbangan atribut fisik untuk kompatibilitas praktis.

Meskipun pencapaian teknis mengesankan, tantangan tetap ada dalam keragaman dan ketersediaan dataset, efisiensi komputasi untuk aplikasi real-time, mitigasi keadilan dan bias, pelestarian privasi dalam personalisasi, dan validasi deployment di dunia nyata.

Penelitian masa depan harus fokus pada representasi fashion yang lebih inklusif dan beragam, promosi fashion berkelanjutan, peningkatan pengalaman pengguna dengan teknologi AR/VR, teknik personalisasi yang menjaga privasi, dan rekomendasi yang dapat diinterpretasi dan dijelaskan.

Dengan penelitian dan pengembangan berkelanjutan, sistem rekomendasi fashion memiliki potensi untuk merevolusi pengalaman belanja online, mempromosikan pilihan fashion berkelanjutan, dan mendemokratisasi akses ke keahlian fashion personal.



DAFTAR PUSTAKA

- [1] He, T., & Hu, Y. (2018). FashionNet: Personalized Outfit Recommendation with Deep Neural Network. arXiv preprint arXiv:1810.02443v1.
- [2] Padilla Reyes, L. J., Bonifaz Oviedo, N., Camacho, E. C., & Calderon, J. M. (2021). Adaptable Recommendation System for Outfit Selection with Deep Learning Approach. IFAC PapersOnLine, 54(13), 605-610.
- [3] Khalid, M., Keming, M., & Hussain, T. (2021). Design and implementation of clothing fashion style recommendation system using deep learning. Romanian Journal of Information Technology and Automatic Control, 31(4), 123-136.
- [4] Huang, Y., & Huang, T. (2017). Outfit Recommendation System Based on Deep Learning. 2nd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2017), Advances in Computer Science Research, 74, 170-174.
- [5] Tuinhof, H., Pirker, C., & Haltmeier, M. (2018). Image Based Fashion Product Recommendation with Deep Learning. arXiv preprint arXiv:1805.08694v2.
- [6] Bag, V. V., Patil, M. B., Gaikwad, V. D., Mohare, R., Abhang, S. P., & Antad, S. (2024). Revolutionizing Fashion: Fashion Era's Deep Convolutional Neural Network for Outfit Recommendations. International Journal of Intelligent Systems and Applications in Engineering, 12(15s), 644-650.
- [7] Bhure, B. P., Bansod, P. T., Amgaokar, M. S., Lodiwale, S. P., Orkey, A. P., & Mohod, A. (2021). Fashion Outfit Recommendation Based on Deep Learning Model. International Journal of Scientific Research in Science and Technology, 8(3), 216-225.
- [8] Ito, T., Nakamura, I., Tanaka, S., Sakai, T., Kato, T., Fukazawa, Y., & Yoshimura, T. (2021). Deep Neural Network Incorporating CNN and MF for Item-Based Fashion Recommendation. Lecture Notes in Computer Science, Springer.
- [9] Rathod, D., Tundare, A., Shelke, P., Muneshwar, S., & Dawkhar, S. L. (2024). Fashion Recommendation Using Deep Learning. International Research Journal of Modernization in Engineering Technology and Science, 6(1), 1979-1982.
- [10] Mo, D., Zou, X., & Wong, W. K. (2023). Personalized Fashion Recommendation via Deep Personality Learning. Proceedings of the British Machine Vision Conference (BMVC).