mulut, perbedaan distribusi warna, serta pola frekuensi yang tidak alami [13].

Beberapa arsitektur CNN populer seperti VGG, ResNet, EfficientNet, dan Xception telah banyak digunakan dalam tugas deteksi gambar deepfake maupun deephoax. Secara khusus, arsitektur Xception yang mengadopsi konsep depthwise separable convolution terbukti efektif dalam menangkap korelasi spasial dan kanal secara lebih efisien dibandingkan CNN konvensional [14]. Pendekatan transfer learning pada model-model tersebut memungkinkan pemanfaatan pengetahuan dari dataset berskala besar, seperti ImageNet, sehingga meningkatkan performa deteksi meskipun dataset target memiliki karakteristik yang berbeda.

2.1.2 Karakteristik dan Ancaman Gambar Deephoax

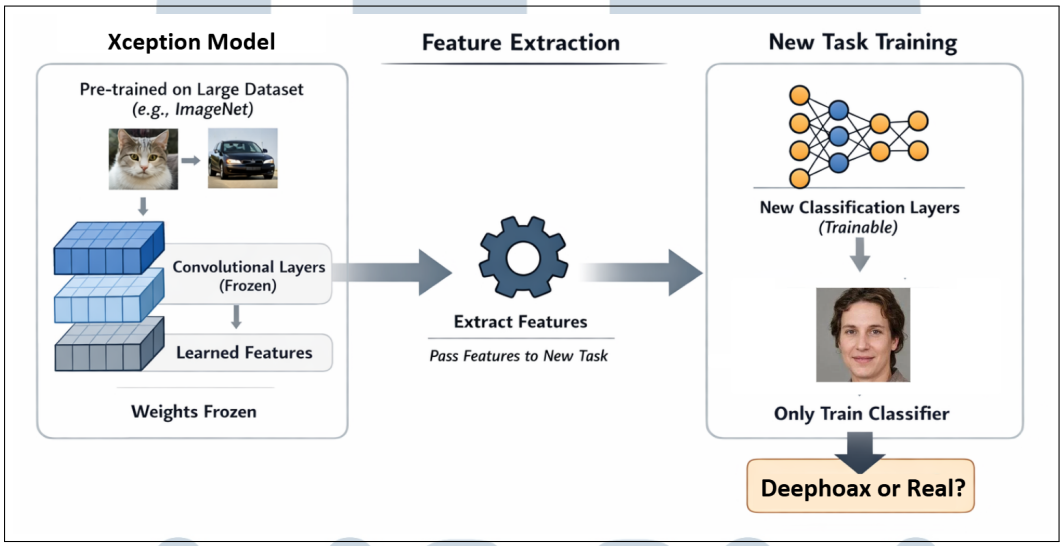
Gambar hasil deephoax memiliki karakteristik visual tertentu yang membedakannya dari gambar asli, meskipun perbedaan tersebut sering kali sulit dikenali oleh manusia. Beberapa karakteristik umum meliputi ketidakkonsistenan tekstur wajah, artefak halus pada area mata, hidung, dan mulut, ketidakseimbangan pencahayaan, serta pola frekuensi gambar yang tidak alami. Karakteristik inilah yang menjadi dasar bagi pendekatan komputasional dalam mendeteksi gambar deephoax melalui analisis fitur visual. Meskipun teknologi deepfake dan deephoax dapat dimanfaatkan untuk tujuan positif seperti hiburan atau industri kreatif, penyalahgunaannya menimbulkan berbagai ancaman serius, antara lain:

1. Penipuan identitas digital, di mana gambar wajah palsu digunakan untuk melewati sistem verifikasi.
2. Penyebaran informasi palsu, yang dapat memanipulasi opini publik dan menurunkan kepercayaan terhadap media digital.
3. Pelanggaran privasi dan reputasi, terutama bagi individu atau figur publik yang wajahnya digunakan tanpa izin dalam konten palsu.

Oleh karena itu, pengembangan metode deteksi gambar deephoax yang andal dan robust menjadi kebutuhan mendesak dalam bidang *computer vision* dan *multimedia forensic*. Metode deteksi yang efektif diharapkan mampu membedakan secara akurat antara gambar asli dan gambar hasil ai generated, meskipun perbedaan visual yang dihasilkan semakin sulit dikenali oleh pengamatan manusia. Selain itu, penelitian di bidang ini juga berperan penting dalam mendukung pengembangan sistem keamanan digital serta menjaga kepercayaan masyarakat terhadap informasi visual yang beredar di era digital.

2.2 Transfer Learning

Transfer learning merupakan pendekatan dalam *machine learning* yang memanfaatkan pengetahuan dari model yang telah dilatih sebelumnya pada dataset berskala besar untuk menyelesaikan permasalahan baru yang memiliki karakteristik serupa. Dalam konteks *computer vision*, transfer learning umumnya dilakukan dengan menggunakan model *pre-trained* yang telah dilatih pada dataset umum seperti ImageNet, kemudian diadaptasi untuk tugas klasifikasi gambar spesifik, termasuk deteksi gambar deephoax [15].



Gambar 2.1. Alur Transfer Learning dengan Pendekatan Feature Extraction

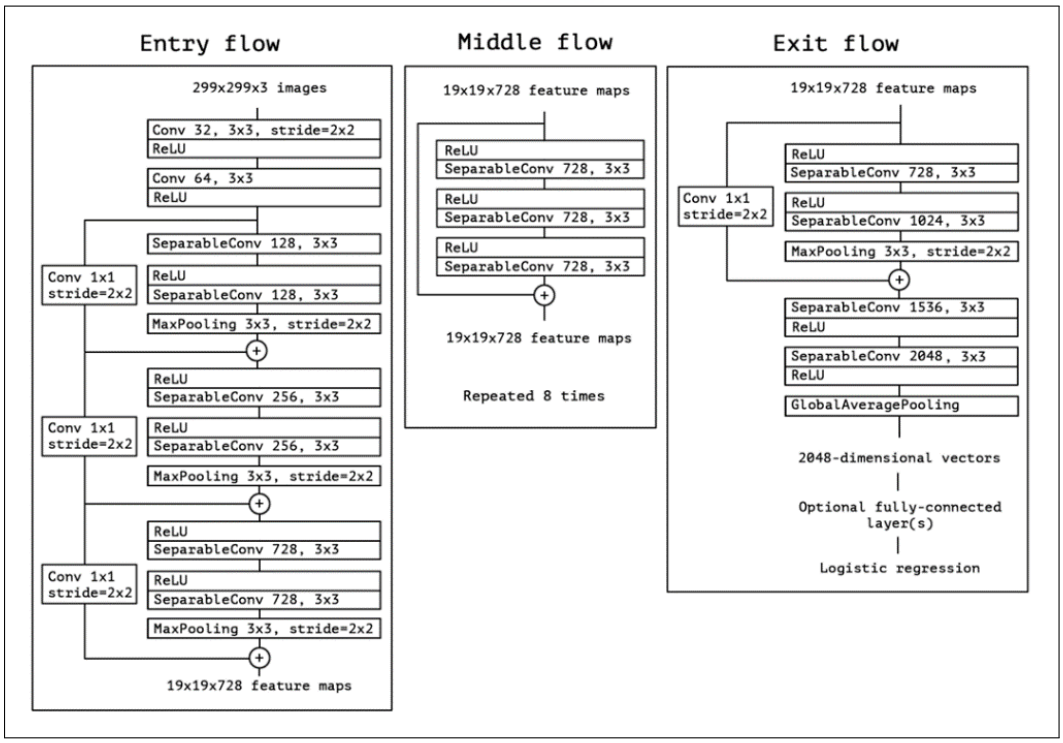
Gambar 2.1 mengilustrasikan proses pemanfaatan model pre-trained sebagai *feature extractor*. Pada pendekatan ini, *convolusional layers* pada model dasar dibekukan (freeze) sehingga bobot hasil pelatihan sebelumnya tetap dipertahankan, sementara proses pelatihan hanya dilakukan pada lapisan klasifikasi di bagian akhir. Pendekatan ini memberikan keuntungan signifikan, terutama ketika dataset yang tersedia relatif terbatas, karena mampu mempercepat konvergensi model, meningkatkan efisiensi training, serta mengurangi risiko overfitting. Dalam penelitian ini, transfer learning digunakan untuk menyesuaikan arsitektur Xception agar dapat mengekstraksi dan membedakan fitur visual penting antara gambar asli dan gambar deephoax secara optimal.

2.3 Xception

Xception (Extreme Inception) merupakan arsitektur deep convolutional neural network yang dikembangkan oleh François Chollet. Xception ini adalah pengembangan dari arsitektur Inception dengan konsep pemisahan penuh antara pemetaan kanal (channel-wise) dan pemetaan spasial (spatial-wise). Xception dirancang untuk meningkatkan efisiensi dan performa model dengan memanfaatkan depthwise separable convolution, sehingga mampu mengekstraksi fitur visual secara lebih efektif dengan jumlah parameter yang relatif lebih sedikit. Dalam penelitian ini, arsitektur Xception digunakan sebagai model feature extractor berbasis transfer learning untuk membedakan karakteristik visual antara gambar asli dan gambar deephoax.

2.3.1 Arsitektur Xception

Arsitektur Xception terdiri dari tiga bagian utama, yaitu entry flow, middle flow, dan exit flow. Struktur ini dirancang untuk mengekstraksi fitur visual secara bertahap, mulai dari fitur dasar hingga fitur tingkat tinggi yang bersifat semantik.



Gambar 2.2. Arsitektur Xception

Sumber: [16]

Gambar 2.2 memperlihatkan alur pemrosesan data dalam jaringan Xception yang dimulai dari input gambar dan diproses melalui tahapan berikut:

1. Entry Flow. Tahap awal yang bertujuan mengekstraksi fitur dasar dari gambar, seperti tepi dan tekstur awal. Pada tahap ini dilakukan beberapa operasi konvolusi dan depthwise separable convolution yang disertai dengan batch normalization dan fungsi aktivasi ReLU. Selain itu, residual connection digunakan untuk menjaga aliran gradien selama pelatihan.

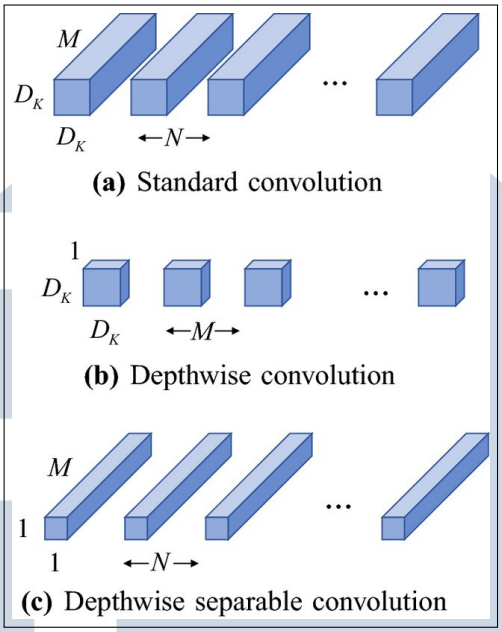
2. Middle Flow. Tahap ini merupakan inti dari arsitektur Xception dan terdiri dari sejumlah blok yang diulang beberapa kali. Setiap blok menggunakan depthwise separable convolution dengan koneksi residual. Middle flow berfungsi untuk memperdalam representasi fitur dan menangkap pola visual kompleks tanpa meningkatkan jumlah parameter secara signifikan.

3. Exit Flow. Tahap akhir yang bertugas mengekstraksi fitur tingkat tinggi sebelum dilakukan klasifikasi. Pada bagian ini, dimensi fitur diperkaya dan kemudian dihubungkan dengan lapisan global average pooling serta fully connected layer untuk menghasilkan keluaran klasifikasi.

Kombinasi ketiga tahap tersebut memungkinkan Xception menghasilkan representasi fitur yang kuat dan diskriminatif, sehingga efektif digunakan dalam tugas klasifikasi gambar berbasis manipulasi visual seperti deteksi deephoax.

2.3.2 Depthwise Separable Convolution

Salah satu karakteristik utama arsitektur Xception adalah penggunaan depthwise separable convolution, yaitu teknik konvolusi yang memisahkan proses ekstraksi fitur spasial dan penggabungan informasi antar kanal. Berbeda dengan standard convolution yang melakukan konvolusi spasial dan kanal secara bersamaan dalam satu operasi, pendekatan ini memecah proses konvolusi menjadi dua tahap utama, yaitu depthwise convolution dan pointwise convolution. Pemisahan ini secara matematis mampu mengurangi jumlah parameter dan beban komputasi secara signifikan, sehingga model menjadi lebih efisien tanpa mengorbankan kemampuan ekstraksi fitur.



Gambar 2.3. Perbedaan Standard Convolution, Depthwise Separable dan Depthwise Separable Convolution

Sumber: [17]

Gambar 2.3 menampilkan perbandingan antara standard convolution, depthwise convolution, dan depthwise separable convolution. Pada standard convolution, satu kernel diterapkan pada seluruh kanal input secara bersamaan untuk menghasilkan satu fitur keluaran. Sebaliknya, pada depthwise convolution, setiap kanal input diproses secara terpisah menggunakan satu kernel, sehingga ekstraksi fitur spasial dilakukan secara independen pada tiap kanal. Selanjutnya, pada depthwise separable convolution, hasil dari depthwise convolution digabungkan menggunakan pointwise convolution dengan kernel berukuran 1×1 untuk mengombinasikan informasi antar kanal.

Pendekatan depthwise separable convolution memungkinkan jaringan untuk mempelajari representasi fitur visual yang lebih spesifik dan efisien, terutama dalam menangkap pola halus pada gambar. Dalam penelitian ini, penerapan arsitektur Xception dengan mekanisme tersebut memberikan keuntungan dalam mendeteksi perbedaan antara gambar asli dan gambar deephoax, khususnya pada tekstur wajah dan artefak visual hasil manipulasi, dengan tetap menjaga efisiensi komputasi selama proses pelatihan dan inferensi.

2.4 Evaluasi Model

Evaluasi model dilakukan untuk menilai sejauh mana model klasifikasi yang dibangun mampu membedakan antara gambar asli (real) dan gambar hasil manipulasi (deephoax). Dalam penelitian ini, evaluasi model difokuskan pada pengukuran kinerja model Xception berbasis transfer learning dalam mendeteksi pola visual yang mengindikasikan adanya manipulasi gambar. Evaluasi dilakukan menggunakan beberapa metrik yang umum digunakan dalam tugas klasifikasi gambar berbasis machine learning. Beberapa metrik yang digunakan dalam evaluasi model meliputi:

1. Accuracy, yaitu metrik yang mengukur proporsi prediksi yang benar terhadap seluruh data uji. Accuracy memberikan gambaran umum performa model, namun kurang representatif pada kasus deteksi deephoax jika terjadi ketidakseimbangan data antar kelas.

2. Loss, yang merepresentasikan tingkat kesalahan model selama proses pelatihan dan validasi. Penurunan nilai loss menunjukkan peningkatan kemampuan model dalam mempelajari pola data, sedangkan perbandingan training dan validation loss digunakan untuk mengidentifikasi overfitting atau underfitting.

3. Confusion Matrix, digunakan untuk membandingkan hasil prediksi model dengan nilai ground truth sehingga dapat diketahui distribusi prediksi benar dan salah pada masing-masing kelas gambar asli dan deephoax.

4. Precision, Recall, dan F1-score, di mana precision mengukur ketepatan prediksi deephoax, recall mengukur kemampuan model mendeteksi seluruh gambar deephoax, dan F1-score merepresentasikan keseimbangan antara precision dan recall dalam evaluasi klasifikasi.

Penggunaan kombinasi metrik accuracy, loss, Confusion Matrix, precision, recall, dan F1-score memungkinkan evaluasi model dilakukan secara komprehensif. Dengan demikian, kekuatan dan kelemahan model Xception dalam mendeteksi gambar deephoax dapat dianalisis secara lebih mendalam serta memungkinkan perbandingan performa dengan pendekatan atau model lain yang dikembangkan pada penelitian sejenis.