

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penelitian pengenalan pola citra berfokus pada pengembangan metode dan algoritma pengenalan pola citra untuk mengidentifikasi pola dan warna pada citra, khususnya pada ikan koi. Dengan karakteristik warna dan pola ikan yang beragam pada tubuhnya, penelitian ini menerapkan arsitektur *machine learning* untuk meningkatkan kinerja identifikasi citra ikan koi.

Ikan koi merupakan salah satu jenis ikan hias yang sangat digemari pengoleksi ikan hias dan populer di seluruh dunia, terutama di Jepang. Sejauh ini penelitian tentang klasifikasi citra ikan koi telah dilakukan menggunakan berbagai teknik pengenalan pola citra, termasuk metode klasifikasi seperti *Support Vektor Machine* (SVM) (Sapphira *et al.* 2020). Penelitian Sapphira mengidentifikasi varietas ikan koi berdasarkan gambar, tetapi Sapphira menyatakan bahwa dataset yang diambil dari internet kualitasnya tidak konsisten. Kondisi gambar yang didapat memiliki perbedaan pantulan cahaya, bayangan, pantulan air, atau riak air yang berbeda.

Penelitian lebih lanjut dilakukan oleh (Cueto *et al.* 2021) untuk mengatasi masalah yang ada pada penelitian sebelumnya, penelitian ini mengimplementasikan *Convolutional Neural Network* (CNN) untuk mengklasifikasikan 15 variasi Koi dengan akurasi sebesar 84%. Dasar pemilihan algoritma CNN menurutnya adalah karena CNN memiliki lapisan yang dilatih sedemikian rupa sehingga mendeteksi pola demi pola yang lebih kompleks dan proses konvolusinya memeriksa seluruh fitur gambar untuk mencapai akurasi yang lebih tinggi dari hasil prediksi. Hasil penelitian tersebut mengatakan CNN mencapai akurasi yang cukup tinggi dalam mengidentifikasi variasi-variasi Koi, namun masih terdapat ruang untuk peningkatan

dalam akurasi klasifikasi. Salah satu cara untuk mencapai ini adalah dengan mengeksplorasi algoritma dan teknik baru.

Terdapat contoh penggunaan arsitektur lainnya yang juga merupakan algoritma *deep learning*. Sebagai contoh adalah arsitektur *Transformer* yang merupakan cikal bakal terbentuknya arsitektur *Vision Transformer*, yaitu pada penelitian oleh (Khan *et al.* 2023) yang membuktikan potensi algoritma tersebut dalam tugas pengenalan pola pada data suara. Penelitian Khan menunjukkan bahwa diantara arsitektur *Deep Learning* yang digunakan dan termasuk di dalamnya arsitektur CNN, *Transformer* mampu mengatasi tantangan dalam pemahaman pola kompleks dalam data dan memberikan hasil yang lebih baik dalam pengenalan pola. Hasil penelitian ini menunjukkan bahwa *Transformer* yang diusulkan dengan mengelola tujuh fitur audio pada semua dataset memberikan kinerja lebih unggul dengan akurasi tertinggi 0,99 untuk ELSDSR, 0,97 untuk CSTR-VCTK, dan 0,99 untuk Ar-DAD. Sedangkan CNN memperoleh akurasi 0,97 untuk ELSDSR, 0,92 untuk CSTR-VCTK, dan 0,97.

Temuan keunggulan arsitektur *Transformer* juga telah diungkapkan pada penelitian (Khan *et al.* 2023), yang mengungkapkan bahwa salah satu kelebihan yang ada pada model *Transformer* adalah memproses urutan data dengan cara yang lebih efisien dibanding arsitektur *deep learning* lainnya. Hal ini memungkinkan *Transformer* memahami pola dan hubungan dalam urutan data yang lebih baik.

Penerapan *Transformer* diawali dalam pemrosesan bahasa alami. *Transformer* merupakan salah satu arsitektur jaringan saraf dalam yang sebagian besar tugasnya mengandalkan mekanisme perhatian diri (*self-attention*). Dalam berbagai tolok ukur visual, *Transformer* memiliki kinerja yang serupa atau lebih baik dibanding jenis jaringan lain seperti jaringan saraf konvolusional (CNN) dan RNN. Berkat kemampuan representasi data yang lebih baik, peneliti mencari cara untuk mengadopsi arsitektur *Transformer* ke dalam konteks tugas-tugas visi komputer, dan akhirnya ditemukanlah arsitektur *Vision Transformer* (ViT) (Han *et al.* 2022).

Arsitektur ViT merupakan adaptasi dari arsitektur *Transformer* yang saat ini menjadi *state-of-the-art* (SOTA) dalam pemrosesan bahasa alami. ViT terdiri dari

beberapa tahap, termasuk *patch embeddings*, *multi-head attention*, dan *multi-layer perceptron* (Figo *et al.* 2020).

Penelitian telah dilakukan untuk mengaplikasikan ViT pada beberapa dataset publik sebagai arsitektur dalam klasifikasi citranya. Contoh penerapan tersebut adalah pada dataset Imagenet1K yang dipakai pada penelitian (Chen *et al.* 2021) yang menggunakan ViT jenis CrossViT-18, kemudian pada dataset kecil seperti CIFAR10/100, CINIC10, SVHN, Tiny-ImageNet, *Aircraft* dan *Cars* berkat penggunaan ViT dengan bantuan *Shifted Patch Tokenization* (SPT) dan *Locality Self-attention* (LSA) (Lee *et al.* 2021), dan dataset *CheXpert* dan *Pediatric Pneumonia* yang digunakan oleh penelitian (Usman *et al.* 2022).

Aplikasi ViT sebagai arsitektur pada dataset pribadi juga sudah digunakan di beberapa penelitian. Sebagai contoh penelitian (Wu *et al.* 2021) yang mengklasifikasikan interpretasi emfisema dengan ViT, kemudian penelitian (Tahyudin *et al.* 2023) yang menggunakan ViT dalam klasifikasi gender berdasarkan citra wajah, dan pada penelitian (Ghali *et al.* 2021) yang menggunakan dua model arsitektur ViT, yaitu TransUNet dan MedT dalam segmentasi gambar pada data kebakaran hutan.

Penelitian yang melibatkan penggunaan arsitektur ViT juga kerap merujuk CNN sebagai model *baseline*-nya dan ViT ditemukan memperoleh hasil kinerja yang lebih baik. Beberapa penelitian yang dimaksud adalah penelitian (Usman *et al.* 2022) yang membuktikan bahwa model dari ViT menunjukkan hasil yang lebih baik dibanding model dari CNN (VGG-16 dan ResNet-50), kemudian penelitian (Chen *et al.* 2021) yang mengusulkan model CrossViT-15 juga menunjukkan hasil dengan performa yang lebih baik dan lebih efisien saat dibandingkan dari CNN keluarga ResNet (termasuk ResNeXt, SENet, ECA-ResNet, dan RegNet) sebagai *baseline*.

Kembali ke uraian sebelumnya, diketahui telah dilakukan penelitian oleh dua penelitian rujukan yang melibatkan penggunaan arsitektur SVM (Sapphira *et al.* 2020) dan arsitektur CNN (Cueto *et al.* 2021) untuk mengidentifikasi 15 variasi koi berdasarkan citra. Menimbang beberapa penemuan bahwa arsitektur *Vision Transformer* (ViT) terbukti lebih unggul dari CNN dalam tugas pengenalan pola citra dan ViT juga masih bagian dari arsitektur *deep learning*, maka penulis mengusulkan

penggunaan arsitektur ViT sebagai *classifier* untuk identifikasi 15 variasi ikan koi berdasarkan citra untuk diteliti performanya dalam mengidentifikasi variasi ikan koi.

Tujuan penelitian ini adalah penerapan arsitektur *Vision Transformer* (ViT) dalam mengidentifikasi variasi ikan koi berdasarkan citra. Dataset penelitian ini dirancang dengan mempertahankan kelas dan jumlah kelas variasi koi, serta jumlah dan pembagian dataset masing-masing kelas variasi dari penelitian rujukan. Oleh karena itu penelitian ini dirancang dengan menggunakan 15 kelas variasi koi yang diantaranya adalah Asagi, Bekko, Doitsu Hariwake, Goromo, Hikarimoyo, Kawarimoyo, Kinginrin, Kohaku, Matsuba, Ogon, Sanke, Showa, Shusui, Tancho, dan Utsurimono, serta jumlah dataset yang dikumpulkan adalah 1500 citra (100 dataset per kelas) Yang dibagi menjadi 80 gambar latih (*train data*) dan 20 gambar uji (*test data*). Oleh karena penelitian sebelumnya belum menambahkan perlakuan augmentasi pada gambar latih, perlakuan augmentasi akan ditambahkan pada gambar latih yang digunakan dan akan menambah jumlah gambar latih.

Perolehan dataset akan dilakukan dengan mengunjungi dan mengunduh dataset dari situs indonesiakoishow.com dan validasi dataset akan dilakukan dengan mengunjungi peternakan & budidaya ikan koi yang bernama “Berkah Koi”, Pusat Peternakan & Budidaya Ikan Koi & Nila.

Proses validasi model ViT yang digunakan dalam penelitian dirancang dengan menggunakan metode *hold-out* (Dewi 2021), yaitu sistem validasi dengan cara membagi dataset menjadi data latih dan data uji. Data latih digunakan untuk melatih model dalam mempelajari pola data citra, sedangkan data uji digunakan untuk menguji akurasi model dalam mengidentifikasi variasi ikan koi. Hasil identifikasi oleh model akan dievaluasi dengan *confusion matrix* dan mengukur metrik performa (Grandini *et al.* 2020). Metrik yang didapat akan digunakan sebagai langkah dalam menyimpulkan kinerja dari arsitektur ViT.

1.2 Rumusan Masalah

Berdasarkan identifikasi masalah yang ada pada latar belakang, maka penulis merumuskan masalah yang akan diteliti untuk disimpulkan hasil penelitiannya pada bab terakhir. Adapun rumusan masalah tersebut adalah:

1. Bagaimana mengklasifikasikan 15 variasi ikan koi dengan menggunakan arsitektur *Vision Transformer*?
2. Bagaimana performa *Vision Transformer* dalam mengklasifikasikan 15 variasi ikan koi dalam pengujian 300 data uji dan performa klasifikasi per kelasnya?
3. Bagaimana performa model dari *Vision Transformer* jika dilakukan teknik augmentasi pada gambar latih?

1.3 Batasan Masalah

Agar pembahasan yang ada pada penelitian ini tidak menyimpang dan sesuai dengan pokok permasalahan, penulis membatasi ruang lingkup dari objek penelitian. Adapun ruang lingkup penelitian ini antara lain adalah:

1. Citra yang diolah dalam penelitian ini adalah sebanyak 1500 citra ikan koi yang terdiri dari 15 kelas variasi ikan koi, diantaranya adalah Asagi, Bekko, Doitsu Hariwake, Goromo, Hikarimoyo, Kawarimoyo, Kinginrin, Kohaku, Matsuba, Ogon, Sanke, Showa, Shusui, Tancho, dan Utsurimono.
2. Dataset citra diunduh dari situs “indonesiakoishow.com” dan telah divalidasi.
3. Gambar latih dan uji memiliki objek koi yang memiliki tampilan sudut pandang atas, posisi badan vertikal, dan posisi kepala berada dibawah. Khusus gambar latih akan diberi perlakuan augmentasi rotasi.
4. File gambar berformat ‘.jpg’ dan diubah menjadi ‘.png’ sebelum diolah oleh model.
5. Pengolahan citra digital menggunakan bahasa pemrograman *Python* dengan *Library PyTorch*.
6. Metode yang digunakan untuk klasifikasi 15 variasi ikan koi ini adalah *Vision Transformer (ViT)*.

1.4 Tujuan Penelitian

Dari rumusan masalah, beberapa tujuan penelitian telah ditetapkan sebagai panduan untuk mencapai hasil penelitian yang diharapkan pada rumusan masalah. Tujuan dari penelitian ini adalah:

1. Membangun model klasifikasi 15 variasi ikan koi dengan menggunakan arsitektur *Vision Transformer*.
2. Mengetahui performa *Vision Transformer* dalam mengklasifikasikan 15 variasi ikan koi dalam pengujian 300 data uji dan performa klasifikasi per kelasnya.
3. Mengetahui performa model dari *Vision Transformer* setelah penambahan teknik augmentasi pada gambar latih.

1.5 Manfaat Penelitian

Penelitian yang baik akan memberikan manfaat bagi banyak orang, objek penelitian, dan sebagai pengayaan literatur pada metode yang diaplikasikan. Manfaat yang ingin diperoleh dari penelitian ini adalah:

1. Mempermudah penghobi koi yang masih awam dalam mengidentifikasi variasi koi berdasarkan gambar.
2. Penelitian ini diharapkan dapat memberikan masukan bagi para peneliti yang tertarik dalam pengembangan teknologi klasifikasi citra koi.
3. Penelitian ini diharapkan dapat menjadi informasi bagi peneliti lainnya dalam memilih arsitektur untuk diaplikasikan pada bidang klasifikasi gambar lainnya.

1.6 Sistematika Penelitian

Sistematika penelitian ditulis agar penelitian lebih mudah dibaca dan dimengerti. Berdasarkan tujuan tersebut, penulis mengklasifikasikan penelitian ini

kedalam enam bab. Adapun sistematika penelitian dalam tesis ini adalah sebagai berikut:

BAB I PENDAHULUAN

Berisi latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

BAB II LANDASAN TEORI

Pada bab ini dijelaskan tentang teori-teori yang mendukung penerapan metode. Teori-teori ini termasuk machine learning, *Vision Transformer* (ViT), Pengolahan citra, dan studi terkait.

BAB III METODOLOGI PENELITIAN

Berisi pembahasan tentang proses yang dilakukan dalam penelitian, proses tersebut dijelaskan mulai dari metode pengumpulan data, analisis dan implementasi, kemudian diakhiri dengan penarikan kesimpulan.

BAB IV ANALISIS DAN PERANCANGAN

Berisi tentang analisis dan perancangan ViT untuk penerapan yang akan digunakan sebagai dasar untuk tahap implementasi selanjutnya.

BAB V IMPLEMENTASI DAN HASIL

Berisi tentang hasil implementasi model ViT pada data. Bab ini juga mencakup pengujian dan pengukuran akurasi model.

BAB VI PENUTUP

Berisikan kesimpulan dari hasil penelitian klasifikasi 15 variasi citra ikan koi menggunakan ViT dan memberikan saran bagi peneliti berikutnya.