

BAB 1 PENDAHULUAN

1.1 Latar Belakang

Segmentasi semantik citra adalah proses mencari sekumpulan *pixel* yang termasuk ke dalam kelompok yang sama. Pemahaman terhadap citra digital menjadi fokus dalam penelitian terkini. Dalam bidang *computer vision*, segmentasi citra adalah salah satu masalah yang tertua dan paling banyak dipelajari. [4]. Adapun semantik menurut KBBI, semantik adalah cabang linguistik yang mempelajari arti/makna yang terkandung pada suatu representasi (dalam hal ini citra perkotaan). Segmentasi semantik merupakan proses tingkat tinggi yang memberikan label terhadap setiap *pixel* pada citra digital yang merepresentasikan suatu objek [1]. Pada aplikasinya untuk keselamatan berkendara, sistem penglihatan digunakan untuk mengenali objek dan kondisi jalan dari citra perkotaan guna mengambil keputusan kendali yang tepat.

Secara umum, proses segmentasi semantik citra terbagi ke dalam dua tahapan, yaitu tahapan ekstraksi fitur dari setiap *pixel* dan tahapan klasifikasi *pixel* ke dalam salah satu *class*. Sebuah fitur adalah sifat unik pada citra yang membedakan antara satu citra dengan yang lainnya untuk menyelesaikan suatu proses komputasi tertentu [1]. Dahulu, ekstraksi fitur dilakukan dengan mengaplikasikan algoritme tertentu. Pendekatan ini disebut juga dengan *hand engineered features*. Beberapa contoh algoritme yang digunakan untuk ekstraksi fitur adalah *Histogram of Oriented Gradients (HOG)*, *Scale-Invariant Feature Transform (SIFT)*, dan *Speeded-Up Robust Features (SURF)* [1]. Tahap selanjutnya adalah menentukan kelas dari masing-masing *pixel*. Algoritme *machine learning* umum digunakan pada tahap ini, sebagai contoh *Support Vector Machine (SVM)* dan *Random Decision Forest (RDF)*. Akan tetapi, pendekatan dalam dua tahapan ini membutuhkan *preprocessing* citra dan *domain knowledge* yang menyeluruh [1]. Permasalahan ini diselesaikan dengan pendekatan berbasis *deep learning*, khususnya *Convolutional Neural Network*. Pendekatan tradisional berbasis *hand engineered features* dan *machine learning* sudah tergantikan dengan pendekatan *deep learning*, yang sudah terbukti lebih efisien dan efektif dengan ketersediaan sumber daya komputasi dan data yang kian meningkat [1]. Pendekatan *deep learning* bekerja dengan cara melatih sebuah *Deep Neural Network*, khususnya *Convolutional Neural Network (CNN)*, untuk melakukan ekstraksi fitur dan klasifikasi secara *end-to-end*. CNN melakukan ekstraksi fitur dengan mengurangi resolusi citra menjadi *feature map (downsampling)* dan mengembalikan resolusi

feature map ke dalam resolusi citra asal (*upsampling*). Literatur yang ada saat ini kebanyakan mengusulkan inovasi arsitektur *neural network*, seperti *Fully Convolutional Network* (FCN) [7], *U-Net*[8], dan *DeepLab*[10].

Topik segmentasi semantik berbasis *deep learning* mulai diteliti secara aktif sejak implementasi *Fully Convolutional Network* (FCN) [7]. Indikator akurasi yang paling penting dalam masalah segmentasi semantik adalah *mean Intersection over Union* (mIoU), yaitu perbandingan jumlah *pixel* yang berhasil diprediksi dengan tepat dengan jumlah semua *pixel*, di mana semakin tinggi mIoU maka semakin baik akurasi prediksi arsitektur tersebut. Berikut ini dijelaskan *state-of-the-art* CNN untuk segmentasi semantik.

Penelitian yang dilakukan Long, et al. [7] merupakan pionir penggunaan *Convolutional Neural Network* untuk segmentasi semantik. Metode yang diberi nama *Fully Convolutional Network* (FCN) ini mengadaptasi arsitektur CNN untuk klasifikasi, yaitu *AlexNet*, *VGGNet*, dan *GoogLeNet* menjadi sebuah *Fully Convolutional Network* dan menggunakan nilai-nilai bobot yang dipelajari dengan melakukan *fine-tuning* untuk proses segmentasi. *Fully convolutional network* mengungguli pendekatan *state-of-the-art* pada saat itu, dengan tingkat akurasi mIoU sebesar 62.7% pada *dataset* PASCAL VOC 2011 dan 62.2% pada *dataset* PASCAL VOC 2012.

Penelitian yang dilakukan oleh Ronneberger, et al. [8] mengusulkan sebuah arsitektur berbasis *encoder decoder U-Net* yang memungkinkan lokalisasi dengan tepat. Penelitian ini menggunakan teknik *data augmentation* untuk menggunakan *dataset* yang tersedia secara efisien. Implementasi arsitektur *U-Net* berhasil memenangkan kompetisi *ISBI cell tracking challenge 2015*, sebuah kompetisi segmentasi citra biomedis, dengan tingkat akurasi mIoU sebesar 92.03% pada *dataset* biomedis PhC-U373 dan 77.56% pada *dataset* DIC-HeLa.

Penelitian yang dilakukan oleh Chen, et al. [10] mengusulkan penggabungan arsitektur *encoder-decoder* dengan *atrous convolution* untuk meningkatkan akurasi prediksi. Arsitektur yang bernama *DeepLab* mampu mengungguli arsitektur lainnya terutama pada prediksi tepian. Saat ini, *DeepLabV3+* [10] adalah iterasi keempat dari *DeepLab* dan merupakan arsitektur dengan tingkat akurasi mIoU terbaik pada *dataset* PASCAL VOC 2012 sebesar 87.8%.

Penelitian yang dilakukan oleh Zhao, et al. [3] mengusulkan arsitektur yang memperhitungkan informasi pada berbagai skala pada *Pyramid Scene Parsing Network* (PSPNet) yang mengekstrak dan menggabungkan informasi konteks

global. Penelitian ini menghasilkan tingkat akurasi 82.6% pada *dataset* PASCAL VOC 2012 *testing set* dan 78.2% pada *dataset* Cityscapes.

Penelitian yang dilakukan oleh Li, et al. melakukan pengujian performa arsitektur *Fully Convolutional Network* (FCN), *U-Net*, dan *DeepLabV3+* untuk melakukan segmentasi semantik citra perkotaan pada *dataset* CamVid [11]. Penelitian ini mengklaim bahwa performa terbaik dimiliki oleh arsitektur FCN, disusul dengan *DeepLabV3+* dan *U-Net* dengan tingkat akurasi mIoU masing masing 74.69%, 73.37%, dan 67.43%.

Berdasarkan penelitian yang sudah ada, penelitian ini menggunakan metode CNN dengan arsitektur *DeepLabV3+* [10] untuk melakukan segmentasi semantik citra perkotaan. Arsitektur tersebut dipilih karena menunjukkan performa *state-of-the-art dataset test* PASCAL VOC 2012 sebesar 89.0%. Objek penelitian adalah *dataset* citra perkotaan Bandung Cityscapes yang dikumpulkan di sekitar daerah Kota Bandung, Jawa Barat, Indonesia. Pengujian dilakukan dengan mengombinasikan nilai *epoch*, *learning rate*, dan jumlah *filter* konvolusi untuk mencari nilai akurasi mean intersection over union yang terbaik.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, penelitian ini membahas beberapa masalah sebagai berikut:

1. Berapa akurasi tertinggi sistem segmentasi semantik dengan metode *convolutional neural network* untuk melakukan segmentasi semantik citra perkotaan Kota Bandung?
2. Bagaimana pengaruh kombinasi nilai *epoch*, *learning rate*, dan jumlah *filter* terhadap akurasi sistem segmentasi semantik citra perkotaan?

1.3 Tujuan Penelitian

Penelitian ini dilakukan dengan tujuan sebagai berikut:

1. Menerapkan metode *convolutional neural network* dengan untuk melakukan segmentasi semantik citra perkotaan di Kota Bandung.
2. Menganalisis pengaruh nilai *epoch*, *learning rate*, dan jumlah *filter* terhadap akurasi sistem segmentasi semantik citra perkotaan.
3. Mengembangkan aplikasi sederhana untuk memfasilitasi proses segmentasi semantik citra perkotaan.

1.4 Batasan Masalah

Dalam penelitian ini, ada beberapa batasan masalah sebagai berikut:

1. Melakukan segmentasi semantik citra perkotaan dengan *dataset* yang dikumpulkan sendiri pada siang hari di Kota Bandung, Jawa Barat, Indonesia.
2. Input citra digital bersifat RGB atau 3-channel
3. Output merupakan prediksi citra yang sudah dengan anotasi setiap *pixel*.

1.5 Kontribusi Penelitian

Kontribusi yang dihasilkan dari penelitian ini adalah:

1. Mengumpulkan *dataset* segmentasi semantik citra perkotaan di kawasan Kota Bandung, Jawa Barat, Indonesia.
2. Menerapkan dan menganalisis metode CNN untuk melakukan segmentasi semantik citra perkotaan.
3. Mencari nilai *epoch*, *learning rate*, dan jumlah *filter* untuk CNN yang paling optimal.

1.6 Metodologi Penelitian

Metodologi penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut.

1. Studi Literatur

Penulisan ini dimulai dengan studi kepustakaan yaitu mengumpulkan bahan-bahan referensi dari buku, artikel, paper, jurnal, makalah mengenai segmentasi semantik citra perkotaan.

2. Pengambilan Sampel

Data sampel yang digunakan adalah *dataset* citra perkotaan yang ditangkap menggunakan kamera ponsel. Setiap *pixel* pada citra memiliki anotasi yang terbagi ke dalam 9 *class*: *bicyclist*, *building*, *car*, *pedestrian*, *road*, *sidewalk*, *sign symbol*, *sky*, dan *tree*

3. Analisis Masalah

Pada tahap ini dilakukan analisis permasalahan, batasan-batasan yang dimiliki, dan kebutuhan yang diperlukan untuk menyelesaikan permasalahan yang dianalisis.

4. Pengujian

Pada tahap ini dilakukan pengujian terhadap aplikasi segmentasi semantik berbasis *convolutional neural network* yang telah dibangun.

5. Evaluasi Metode (Kesimpulan)

Pada tahap ini dilakukan pengumpulan data penelitian dan mengambil suatu konklusi dari penelitian yang sudah dilakukan.

6. Dokumentasi

Pada tahap ini, dilakukan pendokumentasian hasil analisis dan implementasi secara tertulis dalam bentuk laporan skripsi.

1.7 Sistematika Pembahasan

Penelitian ini disusun berdasarkan sistematika penulisan sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, tujuan penelitian, batasan masalah, kontribusi penelitian, serta metode penelitian yang digunakan untuk melakukan segmentasi semantik citra perkotaan.

BAB II LANDASAN TEORI

Bab ini berisi penjelasan dasar teori yang mendukung penelitian ini yang meliputi tinjauan pustaka, tinjauan objek, dan *library* yang digunakan.

BAB III ANALISIS DAN PERANCANGAN

Bab ini berisi analisis algoritme *convolutional neural network* yang digunakan untuk melakukan segmentasi semantik citra perkotaan.

BAB IV IMPLEMENTASI DAN PENGUJIAN

Bab ini berisi implementasi dan pengujian arsitektur *convolutional neural network* yang kemudian diuji dengan *dataset* Bandung Cityscapes. Hasil pengujian kemudian dijelaskan dalam bentuk tabel dan tulisan.

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari penelitian yang telah dilakukan dan saran untuk penelitian di masa mendatang.