

## **Ekstraksi Fitur Bangunan Secara Cepat pada Foto UAV Menggunakan Metode *Deep Residual Neural Network* Berbasis FCN**

Danang Setiaji<sup>1,2</sup>, Harintaka<sup>3</sup>

<sup>1</sup> Mahasiswa Magister Teknik Geomatika, Universitas Gadjah Mada  
Jln. Grafika 2, Mlati, Sleman, Yogyakarta 55281, INDONESIA

<sup>2</sup>Badan Informasi Geospasial (BIG)  
Jl. Raya Jakarta - Bogor Km.46, Cibinong Kab.Bogor Jawa Barat 16911 INDONESIA  
Email: [danang.setiaji@big.go.id](mailto:danang.setiaji@big.go.id)

<sup>3</sup> Dosen Departemen Teknik Geodesi, Fakultas Teknik, Universitas Gadjah Mada  
Jln. Grafika 2, Mlati, Sleman, Yogyakarta 55281, INDONESIA

(Diterima 16 Mei 2019, Disetujui 30 Juni 2019)

### **ABSTRAK**

Dalam beberapa dekade terakhir, teknologi penginderaan jauh menjadi salah satu teknik yang paling bagus untuk akuisisi data pada area luas tanpa perlu melakukan survey langsung di lapangan dengan tingkat kedetailan yang tinggi. Dengan perubahan tutupan lahan yang semakin cepat dan dinamis diperlukan teknologi yang dapat menghasilkan data dengan tingkat kedetailan tinggi, murah dan dapat menghasilkan data dalam waktu yang cepat. Salah satu produk penginderaan jauh tersebut adalah Foto Udara dengan teknologi pesawat tanpa awak (UAV). Teknologi ini mampu menghasilkan *output* data spasial detail dengan biaya relatif murah, cepat dan terkini. Salah 1 unsur yang cepat berubah di wilayah perkotaan adalah pada penambahan objek bangunan. Dalam pembuatan bangunan tersebut harus dilengkapi dengan syarat keterangan ijin mendirikan bangunan (IMB) dari instansi terkait. Ketentuan ijin ini tidak terlepas dari aturan RDTR dan PZ yang berlaku serta hanya boleh didirikan pada zonasi yang diijinkan. Dengan kebutuhan data informasi fitur bangunan yang cukup banyak dan mendesak, diperlukan metode ekstraksi yang cepat juga. Metode yang saat ini berkembang pesat adalah ekstraksi otomatis menggunakan teknologi *deep learning*. Metode yang digunakan adalah *Fully Convolutional Network* (FCN) dengan arsitektur *Residual Neural network* (ResNet). Dalam penelitian ini akurasi ekstraksi pada fitur bangunan dapat mencapai 95% dan dapat melakukan ekstraksi pada satu NLP skala 1 : 5.000 dengan cepat. Metode ini dapat menghasilkan ekstraksi pada image seperti foto udara dengan cepat dan akurat.

**Kata kunci - Penginderaan Jauh, Foto udara, UAV, Segmentasi, Deep Learning, Fully Convolutional Network, Residual Neural network**

### **ABSTRACT**

*In the past few decades, remote sensing technology has become one of the best techniques for data acquisition in large areas without the need to direct surveys on the field and produce high levels of detail. With changes in land cover that are increasingly fast and dynamic, technology that can produce data with a high level of detail, cheap and can produce data in a fast time is needed. One example of remote sensing product is Aerial Photography with unmanned aircraft technology (UAV). This technology is able to produce detailed spatial data output at relatively low cost, fast and up to date. One of the fast-changing elements in urban areas is the addition of building objects. In making the real building at some place, it must need building permit (IMB) information from the relevant government agency. These permit conditions cannot be separated from the applicable spatial detail plan and zoning map rules and may only be established on allowable zoning. With the need for data on building features that are quite numerous and urgent, fast extraction methods are needed too. The method that is currently developing rapidly is automatic extraction using deep learning technology. The method used is Fully Convolutional Network (FCN) with Residual Neural network (ResNet) architecture. In this study, the extraction accuracy in building features can reach 95% and can extract at one map sheet at scale of 1: 5,000 quickly. This method can produce extractions on images such as aerial photos quickly and accurately.*

**Keywords – Remote Sensing, Aerial Image, UAV, Segmentation, Deep Learning, Fully Convolutional Network, Residual Neural network**

## 1. PENDAHULUAN

Dalam beberapa dekade terakhir, teknologi penginderaan jauh menjadi salah satu teknik yang paling bagus untuk akuisisi data pada area luas tanpa perlu melakukan survey langsung di lapangan. Peta tutupan lahan merupakan salah 1 produk utama yang dihasilkan dari analisis citra menggunakan teknologi penginderaan jauh (Chuvieco, 2016). Penggunaan data penginderaan jauh untuk pemetaan tutupan lahan memerlukan serangkaian alat untuk memperoleh informasi yang akurat. Oleh karena itu, sejumlah penelitian telah dilakukan untuk mengatasi masalah yang mempengaruhi keakuratan proses interpretasi dan untuk menemukan alat yang sesuai yang dapat mengekstraksi lebih lengkap informasi tutupan lahan yang terkandung dalam citra dibandingkan teknik konvensional (Kassaye, 2006).

Teknologi penginderaan jauh terkini, dapat menghasilkan data tutupan lahan dengan tingkat kedetailan yang tinggi. Dengan perubahan tutupan lahan yang semakin cepat dan dinamis diperlukan teknologi yang dapat menghasilkan data dengan tingkat kedetailan tinggi, murah dan dapat menghasilkan data dalam waktu yang cepat. Salah satu produk penginderaan jauh tersebut adalah Foto Udara dengan teknologi pesawat tanpa awak (UAV). Teknologi ini mampu menghasilkan *output* data spasial detail dengan biaya relatif murah, dengan data luaran berupa mozaic orthophoto yang dapat menghasilkan data semantik dan Digital Terrain Model (DTM) yang dapat menggambarkan relief kenampakan suatu wilayah dengan detail. Untuk saat ini data penginderaan jauh yang diperoleh menggunakan UAV dapat digunakan untuk menghasilkan data tutupan lahan terkini secara efektif pada luasan maksimal sampai 100 km<sup>2</sup> untuk sekali proses pengolahan data. Beberapa penggunaan data tutupan lahan tersebut adalah untuk pembuatan peta rencana detail tata ruang, proses deteksi perubahan penggunaan lahan pada skala detail, identifikasi data jalan untuk perencanaan strategis perkotaan (Mnih dan Hinton, 2010), deteksi estimasi populasi (Torgerson dkk., 2011) dan deteksi kendaraan untuk pemantauan lalu lintas (Gleason dkk., 2011).

Salah 1 unsur yang cepat berubah di wilayah perkotaan adalah pada penambahan objek bangunan. Penambahan bangunan ini dilakukan untuk berbagai fungsi seperti bangunan tempat tinggal, peternakan perdagangan dan jasa serta industri. Dalam pembuatan bangunan tersebut harus dilengkapi dengan syarat keterangan ijin mendirikan bangunan (IMB) dari instansi terkait. Ketentuan ijin ini tidak terlepas dari aturan RDTR dan PZ yang berlaku. Pendirian bangunan tertentu hanya boleh

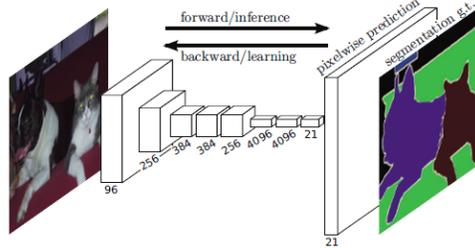
dibangun pada kawasan atau zonasi tertentu saja. Dengan memperhatikan kondisi geografis wilayah Indonesia dan banyaknya RDTR yang harus dipantau dan dievaluasi, diperlukan data penginderaan jauh yang cukup akurat, murah dan dapat diperoleh dengan cepat. Data fitur bangunan ini diperoleh dari ekstraksi citra satelit maupun foto udara. Interpretasi secara interaktif yang dilakukan oleh operator manusia secara visual masih merupakan pendekatan utama untuk klasifikasi digital informasi dari citra penginderaan jauh. Akurasi interpretasi yang dihasilkan tergantung pada keterampilan dan pengalaman dari operator. Sehingga dapat terjadi inkonsistensi pada data yang dihasilkan oleh operator yang berbeda. Interpretasi secara interaktif ini membutuhkan waktu dan sumber daya manusia yang semakin banyak apabila data yang diolah semakin besar (Bergado, 2016).

Ekstraksi secara otomatis saat ini dapat dilakukan dengan pendekatan machine learning. Dalam 5 tahun terakhir, pendekatan machine learning telah menjadi populer dan berkembang pesat untuk proses deteksi, segmentasi dan pengenalan objek dalam suatu gambar atau foto (LeCun dkk., 2015). Metode *machine learning* terbaru yang terkenal adalah pendekatan *deep learning*. *Deep learning* mengacu pada jaringan saraf tiruan yang terdiri dari banyak lapisan neuron buatan yang disusun secara hierarki untuk memproses informasi. Dengan perkembangan perangkat keras komputer saat ini, salah 1 bagian metode *deep learning* yang dianggap paling berhasil adalah menggunakan tipe jaringan yang disebut deep convolutional neural network atau dikenal CNN (Uba, 2016). Arsitektur CNN dapat bekerja secara efektif untuk proses pembelajaran pengenalan objek yang salah 1 fungsinya untuk klasifikasi citra atau foto udara. Penggunaan data foto UAV untuk ekstraksi data tutupan lahan salah satunya pada unsur bangunan dengan metode *deep learning* diharapkan dapat menghasilkan data tutupan lahan yang akurat dan efisien.

## 2. Fully Convolutional Network (FCN)

Ekstraksi fitur pada *image* data raster salah satunya pada foto udara dikenal juga dengan segmentasi semantik. Proses ini menggunakan pendekatan *Fully Convolutional Network* (FCN) yang bekerja dengan mengadaptasi dan modifikasi model rekognisi CNN sehingga cocok digunakan untuk proses segmentasi. Tantangan untuk melakukan modifikasi dari model rekognisi yaitu model itu menghasilkan *output* dalam 1D (hanya berupa probabilitas untuk setiap label) sedangkan model segmentasi menghasilkan *output* 3D (vektor probabilitas untuk setiap pixel). Modifikasi ini dilakukan dengan menghapus layer akhir fully-

connected pada arsitektur CNN dapat menghasilkan model “fully convolutional” yang menghasilkan *output* dalam 3D. Hasil akhir dari metode FCN akan menghasilkan resolusi *output* yang sama dengan resolusi foto input nya (Long dkk., 2015).

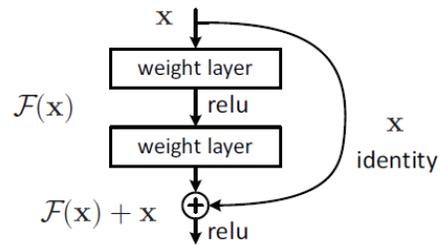


**Gambar 1.** Model arsitektur FCN untuk melakukan segmentasi semantik dengan melakukan prediksi setiap pixel pada *image* (Long dkk., 2015)

**2.1 Residual Neural network (ResNet)**

FCN pada awalnya ditujukan dengan adaptasi dari model rekognisi VGG, tetapi model ini dapat digunakan juga untuk mengadaptasi model rekognisi seperti residual neural network (ResNets). Jaringan ini dikembangkan oleh peneliti dari Microsoft research. Model jaringan ini memperoleh peringkat pertama dalam kompetisi klasifikasi *image* ILSVRC pada tahun 2015 dengan akurasi 96,43% . Jaringan yang digunakan pada model ini memiliki 152 lapisan, jauh lebih dalam dibandingkan model VGG. Model resnet memperkenalkan blok baru yang dikenal dengan blok residual. Karakteristik Jaringan saraf residual memanfaatkan koneksi skip, atau jalan pintas untuk melompati beberapa layer. Sehingga layer – layer tersebut dapat menyalin inputnya pada layer

selanjutnya. Model-model khas ResNet ini diimplementasikan dengan lompatan satu layer (He dkk., 2016).



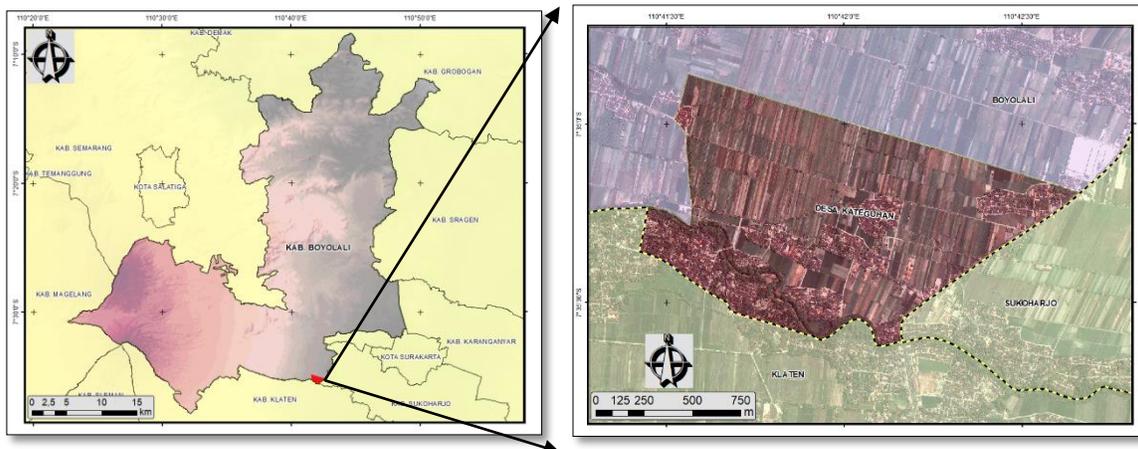
**Gambar 2.** Struktur modul Resnet dengan fungsi skip untuk menghasilkan blok residual (Hedkk., 2016)

Motivasi untuk melakukan skip/melompati 1 lapisan adalah untuk menghindari masalah hilangnya gradien, dengan menggunakan kembali aktivasi dari lapisan sebelumnya sampai lapisan yang berdekatan dapat mempelajari bobotnya.

**3. Metodologi**

**3.1 Bahan Penelitian**

Penelitian ini dilakukan di area Desa Kateguhan, Kabupaten Boyolali, Provinsi Jawa Tengah. Desa ini dilewati oleh jalan nasional yang menghubungkan Kota Yogyakarta dengan Kota Surakarta dan Kota Semarang. Daerah ini merupakan dataran rendah dengan topografi yang datar. Tutupan lahan pada wilayah ini terdiri dari sawah irigasi dan area pemukiman padat dengan bentuk bangunan beraneka ragam serta rata – rata ketinggian bangunan 4 m atau bangunan dengan 1 lantai. Lokasi penelitian ditunjukkan pada gambar 3.



**Gambar 3.** Lokasi Penelitian

Bahan yang digunakan pada penelitian ini adalah orthomosaik foto udara yang diakuisisi pada tahun 2018. Saluran yang dihasilkan adalah red, green dan blue dengan resolusi spasial 15 cm.

**3.2 Peralatan**

Perangkat yang digunakan adalah seperangkat laptop yang menggunakan sistem operasi MS-windows 10 64 bit dengan spesifikasi perangkat keras berupa processor intel i7 2.6 GHz, Ram 16 GB DDR4, ssd 512 GB, VGA GeForce GTX 970M 3 GB GDDR5. Perangkat lunak yang digunakan pada pelaksanaan adalah agisoft metashape, arcgis arcmap, global mapper, notepad++ dan Python anaconda 3 dilengkapi dengan library Tensorflow, Keras, Numpy dan OpenCV. Untuk penulisan laporan menggunakan MS Office 2013.

**3.3 Langkah Kerja**

Secara umum pelaksanaan penelitian ini dimulai dari tahap akuisisi data foto udara dan koreksi geometris. Foto yang telah terkoreksi dilakukan labelling untuk menghasilkan *ground truth* pada beberapa lokasi yang akan digunakan sebagai sampel untuk digunakan sebagai data *training* dan data test. Dengan menggunakan data yang telah siap ini, akan dilakukan desain jaringan dengan arsitektur ResNet50 FCN untuk melakukan segmentasi semantik. Pada penelitian ini, segmentasi dilakukan untuk menghasilkan ekstraksi fitur bangunan. Hasil dari segmentasi semantik secara otomatis ini akan dilakukan pengujian dibandingkan dengan data digitasi *ground truth* secara interaktif.

1) Akuisisi dan Pengolahan Data Foto Udara

Sumber data yang digunakan dalam penelitian ini diperoleh dari hasil akuisisi foto udara yang direkam menggunakan wahana multirotor UAV. Data tersebut diolah menjadi foto orthomosaik menggunakan *metode Structure from motion* (SfM). Data foto orthomosaic yang diolah telah memiliki posisi koordinat mendekati sebenarnya di lapangan. Agar posisi tersebut lebih akurat dilakukan transformasi 2D pada koordinat x dan y dengan menggunakan titik koordinat tanah dari citra satelit tegak resolusi tinggi (CSRT) sebagai referensi, sedangkan data ketinggian menggunakan data posisi absolut dari wahana UAV. Foto orthomosaic yang telah diolah ini berada pada tingkat ketelitian 1 : 5000 untuk posisi 2D.

2) Pemotongan data Foto untuk *training*

Pengolahan pada tahap 1 menghasilkan 1 *scene* data foto orthomosaic dari penggabungan 1109 foto hasil akuisisi menggunakan wahana UAV. Data ini dipotong pada beberapa lokasi untuk digunakan sebagai data *training* sampel (pembelajaran) pada program dengan ukuran setiap potongan sebesar 256 x 256 pixel. Ukuran tersebut disesuaikan dengan kemampuan maksimal yang dapat dicapai pada perangkat keras dan arsitektur jaringan yang digunakan. Data *training* yang dihasilkan dari pemotongan ini sejumlah 215 sampel data.

3) Perancangan jaringan

Jaringan yang digunakan pada penelitian ini menggunakan referensi dari arsitektur ResNet50. Model jaringan ini terdiri dari 50 lapisan jaringan residual. Beberapa jenis arsitektur ResNet beserta dengan ukuran layernya ditampilkan dalam tabel 1 berikut.

**Tabel 1.** setiap blok ResNet terdiri dari 2 layer (Digunakan di jaringan kecil seperti ResNet 18, 34) atau 3 layer (ResNet 50, 101, 152) (Hedkk., 2016).

| layer name | output size | 18-layer  | 34-layer  | 50-layer  | 101-layer  | 152-layer  |
|------------|-------------|---|---|---|--|--|
| conv1      | 112x112     | 7x7, 64, stride 2   |   |   |  |  |
|            |             | 3x3 max pool, stride 2  |   |   |  |  |
| conv2_x    | 56x56       | $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$   | $\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$   | $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$    | $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$     | $\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$     |
| conv3_x    | 28x28       | $\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$ | $\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$ | $\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$  | $\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$   | $\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$   |
| conv4_x    | 14x14       | $\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$ | $\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$ | $\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$ | $\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$ | $\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$ |
| conv5_x    | 7x7         | $\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$ | $\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 8$ | $\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$ | $\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$  | $\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$  |
|            | 1x1         | average pool, 1000-d fc, softmax  |   |   |  |  |
| FLOPs      |             | 1.8x10 <sup>9</sup>   | 3.6x10 <sup>9</sup>   | 3.8x10 <sup>9</sup>   | 7.6x10 <sup>9</sup>  | 11.3x10 <sup>9</sup>   |

4) Evaluasi Akurasi Jaringan

Dilakukan dengan pengujian terhadap data *ground truth* data sampel *training* dan data tes. Evaluasi ini dilakukan untuk mengukur akurasi jaringan yang dibuat pada data *training*. Data yang dihasilkan dari proses segmentasi otomatis dari model jaringan yang dibuat, dibandingkan dengan data *ground truth* dari digitasi interaktif pada objek – objek yang sama. Untuk mendeteksi akurasi yang dihasilkan, digunakan indeks IoU (*Intersection over Union*) atau indeks *Jaccard*. Bagian yang saling memotong (*intersect*) antara data prediksi dengan *ground truth* akan dibandingkan dengan bagian yang bergabung (*union*) (Rezatofighi dkk., 2019). Perhitungan indeks IoU antara area A dan B dapat dirumuskan sebagai berikut :

$$IoU(A, B) = \frac{Area\ of\ Overlap\ (A \cap B)}{Area\ of\ Union\ (A \cup B)} \quad (1)$$

4. Hasil dan Pembahasan

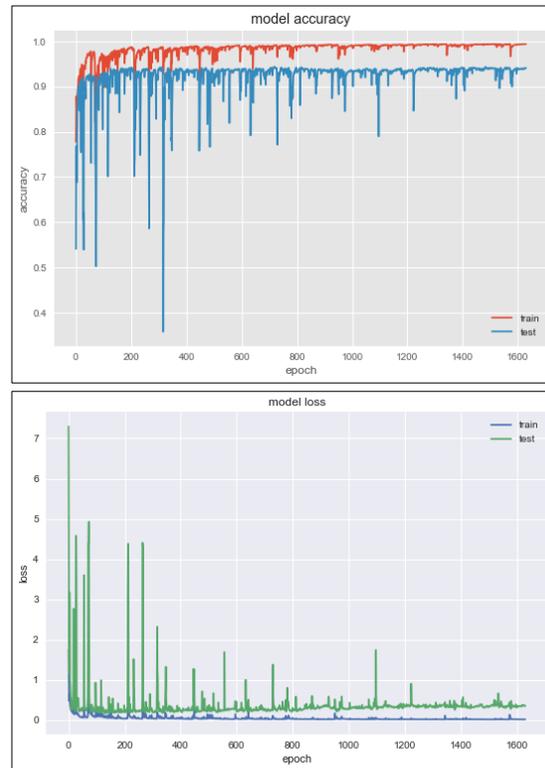
4.1 Performa pada Model Jaringan Resnet

Proses *training* pada 215 sampel data memerlukan waktu selama 40 jam 13 menit sampai menghasilkan akurasi yang diharapkan. Dari sejumlah 215 sampel, sebanyak 10% data sampel atau sejumlah 22 sampel digunakan sebagai data validasi dan 193 sampel sebagai data *training*. Kedua jenis data tersebut akan digunakan dalam pembelajaran jaringan. Performa dari pembelajaran model jaringan dapat ditampilkan pada tabel 2 berikut.

Tabel 2. Tabel performa pelatihan model jaringan ResNet50 (akurasi dalam persen)

| Epoch | Waktu (s) | Training Loss (%) | Training Accuracy (%) | Validation Loss (%) | Validation Accuracy (%) |
|-------|-----------|-------------------|-----------------------|---------------------|-------------------------|
| 1     | 89        | 174.27            | 77.77                 | 729.62              | 54.11                   |
| 150   | 13350     | 5.30              | 97.88                 | 19.26               | 93.55                   |
| 300   | 26700     | 4.14              | 98.33                 | 20.02               | 93.72                   |
| 450   | 40050     | 11.65             | 95.65                 | 52.61               | 84.70                   |
| 600   | 53400     | 6.24              | 98.51                 | 14.91               | 96.08                   |
| 750   | 66750     | 3.85              | 98.43                 | 16.85               | 95.88                   |
| 900   | 80100     | 5.88              | 97.67                 | 14.43               | 95.77                   |
| 1050  | 93450     | 3.34              | 98.64                 | 24.50               | 95.86                   |
| 1200  | 106800    | 3.44              | 98.58                 | 4.47                | 98.15                   |
| 1350  | 120150    | 3.77              | 98.47                 | 7.36                | 97.35                   |
| 1500  | 133500    | 7.83              | 97.26                 | 7.04                | 97.35                   |
| 1627  | 144803    | 2.58              | 98.95                 | 7.19                | 97.69                   |

Dari tabel di atas dapat digambarkan grafik akurasi dan kesalahan dari pembelajaran model jaringan dalam gambar 4 berikut.



Gambar 4. Grafik akurasi dan kesalahan pada model ResNet50

Berdasarkan tabel dan grafik di atas, proses pembelajaran pada model jaringan mengalami peningkatan akurasi yang cukup signifikan antara epoch 1 – 150. Sedangkan epoch selanjutnya akurasi jaringan meningkat secara perlahan dengan *training loss* yang semakin berkurang. Terdapat 2 kolom akurasi yang harus diperhatikan yaitu *training accuracy* dan *validation accuracy*. Kedua nilai tersebut harus memiliki akurasi yang seimbang sehingga hasil ekstraksi bangunan tidak hanya memiliki akurasi yang tinggi pada area yang sama dengan data sampel, tetapi juga memiliki akurasi yang bagus pada area lain. Dari tabel di atas akurasi yang terbaik pada *training* dan validasi, didapatkan pada epoch ke 1627 yaitu sebesar 98,95% untuk akurasi *training* dan 97,69% untuk akurasi data validasi.

4.2 Hasil Ekstraksi dan Akurasi

4.2.1 Validasi dengan Data Sampel Training

*Image* foto pada data ini digunakan juga sebagai data input pada proses pelatihan. Proses ini dilakukan dengan membandingkan antara data *ground truth* dengan ekstraksi secara otomatis pada data *training* sampel. Foto pada data ini digunakan juga pada proses pelatihan. Data yang diuji

berdimensi 256 x 256 pixel pada foto beresolusi 15 cm. Hasil validasi ini ditunjukkan pada gambar 5.

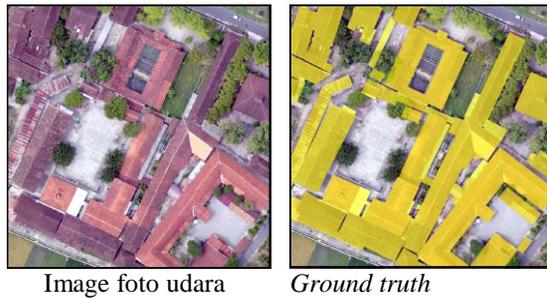
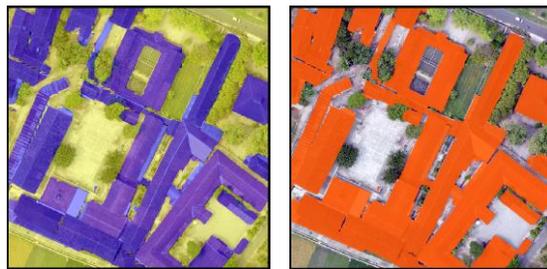


Image foto udara

Ground truth



Hasil segmentasi

Hasil ekstraksi



- Hasil Prediksi yang benar
  - Seharusnya bangunan
  - Seharusnya non – bangunan
- Akurasi ekstraksi data sampel *training*  
 Bangunan : 95.313%  
 Non – Bangunan : 94.726%

**Gambar 5.** Validasi pada data sampel *training* dibandingkan dengan data *ground truth*

Perbandingan hasil ekstraksi otomatis dengan data *ground truth* menghasilkan akurasi sebesar 95.

313%. Akurasi ini dihitung menggunakan indeks IoU. Berdasarkan tampilan visualnya, data yang diuji sudah menunjukkan hasil yang baik. Sebagian besar unsur bangunan sudah diprediksi dengan benar. Terdapat beberapa objek yang seharusnya non-bangunan tetapi diprediksi sebagai bangunan. Hal ini terjadi karena objek tersebut memiliki karakteristik yang hampir mirip dengan unsur bangunan seperti pada permukaan tanah yang di tutup dengan beton, pola dan warna pada beton hampir sama dengan atap berwarna abu – abu atau putih.

**4.2.2 Validasi dengan Data Test**

*Image* foto pada data ini tidak digunakan sebagai data input pada proses pelatihan, sehingga merupakan data lain yang tidak berkaitan (independen). Proses ini dilakukan dengan membandingkan antara data *ground truth* dengan ekstraksi secara otomatis pada data *training* sampel. Foto pada data ini digunakan juga pada proses pelatihan. Data yang diuji berdimensi 256 x 256 pixel pada foto beresolusi 15 cm. Hasil validasi ini ditunjukkan pada gambar 6.



Image foto udara

Ground truth



Hasil segmentasi

Hasil ekstraksi



- Hasil Prediksi yang benar
  - Seharusnya bangunan
  - Seharusnya non – bangunan
- Akurasi ekstraksi data test  
 Bangunan : 93.265%  
 Non – Bangunan : 96.538%

**Gambar 6.** Validasi pada data *test* dibandingkan dengan data *ground truth*

Perbandingan hasil ekstraksi otomatis dengan data *ground truth* menghasilkan akurasi sebesar 93.265%. Akurasi ini dihitung menggunakan indeks IoU. Berdasarkan tampilan visualnya, data yang diuji sudah menunjukkan hasil yang baik walaupun akurasinya lebih rendah dibandingkan validasi pada data sampel *training*. Hal ini terjadi karena data yang digunakan merupakan data independen yang tidak digunakan pada saat proses pelatihan jaringan, tetapi sebagian besar unsur bangunan sudah diprediksi dengan benar. Terdapat beberapa objek yang seharusnya non-bangunan tetapi diprediksi sebagai bangunan. Hal ini terjadi karena objek tersebut memiliki karakteristik yang hampir mirip dengan unsur bangunan, seperti pada jembatan yang permukaannya berupa beton. Objek ini memiliki pola dan warna yang hampir sama dengan atap bangunan.

**4.3 Uji Kecepatan Ekstraksi pada 1 NLP peta 1 : 5.000**

Uji ini dilakukan untuk membandingkan kecepatan ekstraksi secara otomatis dibandingkan dengan digitasi secara interaktif pada ukuran 1 NLP Peta RBI 1 : 5.000. Satu NLP pada skala 1 : 5.000 berdimensi 75” x 75 “ atau ekuivalen dengan 5,38

kilometer persegi. Untuk melakukan ekstraksi secara otomatis membutuhkan waktu selama 47.4 detik pada proses segmentasi, 1 menit 8 detik untuk proses penyimpanan dan 8,68 detik untuk melakukan konversi dari data raster menjadi vektor, sehingga membutuhkan waktu keseluruhan selama 2 menit 4,08 detik. Dengan demikian proses ekstraksi bangunan secara otomatis pada 1 NLP membutuhkan waktu jauh lebih cepat dibandingkan digitasi secara interaktif. Proses ekstraksi secara otomatis pada 1 NLP dapat dilihat pada gambar 7 berikut.

segmentasi

```

%%time
#PREDICT
print('start to predict..')
valx = model.predict(z_test)
print('done')

start to predict..
done
Wall time: 47.4 s
        
```

saving

```

convert_visualize_pixels(prediction)
plt.savefig(folder_testing_gt +'boyXNLP2.tif')
plt.show()

start to stich together
done

Wall time: 1min 8s
        
```

Ekstraksi raster ke polygon

Raster to Polygon

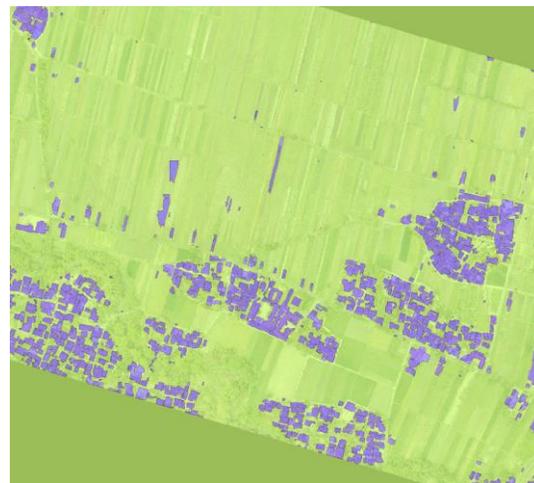
Completed Close

[<< Details](#)

Close this dialog when completed successfully

```

Start Time: Sat May 11
17:34:38 2019
Succeeded at Sat May 11
17:34:46 2019 (Elapsed Time:
8,68 seconds)
                
```



**Gambar 7.** Kecepatan ekstraksi pada 1 NLP secara otomatis

## 5. KESIMPULAN DAN SARAN

Pendekatan metode *deep learning* dengan metode FCN untuk melakukan ekstraksi fitur bangunan dan penggunaan data foto udara UAV mampu memberikan hasil yang akurat dan cepat. Model jaringan ini dirancang menggunakan pre-trained dari ResNet50 yang menghasilkan *output* segmentasi semantik pada kelas bangunan dan non-bangunan. Pembelajaran jaringan dengan model ini menghasilkan akurasi sebesar 98,98%. Sedangkan akurasi segmentasi bangunan yang dihasilkan dapat mencapai 95%. Hasil ekstraksi secara visual memberikan akurasi yang baik dan mendekati bentuk aslinya. Hal ini terjadi disamping penggunaan metode pembelajaran yang tepat, didukung juga oleh kualitas foto udara yang bagus dan beresolusi sangat tinggi. Beberapa kekurangan yang masih terlihat adalah masih terdapat fitur non-bangunan yang di deteksi sebagai bangunan karena karakteristik pola dan warna yang mirip dengan bangunan. Bentuk segmentasi pada beberapa bangunan masih kurang lurus dan tegas. Dari segi kecepatan, metode ekstraksi secara otomatis jauh lebih cepat dibandingkan dengan digitasi secara interaktif. Hal ini sangat bermanfaat terutama untuk area yang luas dan data yang besar.

Untuk meningkatkan akurasi dalam deteksi dan bentuk bangunan agar sesuai dengan aslinya, perlu dilakukan *tuning*/penyempurnaan dengan metode lain sehingga ekstraksi yang dihasilkan menjadi lebih realistis. Untuk meningkatkan akurasi dan kecepatan dari pembelajaran jaringan, memerlukan dukungan penyempurnaan dengan optimalisasi pada perangkat lunak sehingga penggunaan memory pada PC/Notebook menjadi lebih efisien, serta diperlukan dukungan peningkatan perangkat keras sehingga dapat menangani data *training* yang lebih banyak dan tambahan opsi layer lain seperti saluran IR (*infra-red*) atau data ketinggian (DSM).

### UCAPAN TERIMA KASIH

Terimakasih ditujukan kepada Badan Informasi Geospasial dan Pemerintah Kabupaten Boyolali yang memfasilitasi dan memberikan kesempatan untuk melakukan penelitian ini. Peneliti memberikan ucapan terima kasih juga kepada Muhammad Hasannudin Yusa dan Aldino Rizaldy yang telah bersedia bertukar pikiran serta semua pihak yang ikut membantu dan memberikan dukungan dalam proses penelitian ini.

### DAFTAR PUSTAKA

Bergado, J. (2016). A Deep Feature Learning Approach to Urban Scene Classification.

Unpublished MSc. Thesis, University of Twente, Enschede, The Netherlands.

- Chuvieco, E. (2016). *Fundamentals of satellite remote sensing: An environmental approach*: CRC press.
- Gleason, J., Nefian, A. V., Bouyssou, X., Fong, T., dan Bebis, G. (2011). *Vehicle detection from aerial imagery*. Paper presented at the Robotics and Automation (ICRA), 2011 IEEE International Conference on.
- He, K., Zhang, X., Ren, S., dan Sun, J. (2016). *Deep residual learning for image recognition*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Kassaye, R. H. (2006). *Suitability of Markov random field based method for super resolution land cover mapping*.
- LeCun, Y., Bengio, Y., dan Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.
- Long, J., Shelhamer, E., dan Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Mnih, V., dan Hinton, G. E. (2010). *Learning to detect roads in high-resolution aerial images*. Paper presented at the European Conference on Computer Vision.
- Rezatofghi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., dan Savarese, S. (2019). Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. *arXiv preprint arXiv:1902.09630*.
- Torgerson, D. G., Ampleford, E. J., Chiu, G. Y., Gauderman, W. J., Gignoux, C. R., Graves, P. E., . . . Hancock, D. B. (2011). Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nature genetics*, 43(9), 887.
- Uba, N. K. (2016). *Land use and land cover classification using deep learning techniques*: Arizona State University.