Multilabel Image annotation Menggunakan Metode *Convolutional Neural Network*

Naufal Dzaky Anwari[1], Anditya Arifianto,S.T.,M.T.[2], Jondri,S.Si, M.Si[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]naufalyai@students.telkomuniversity.ac.id, [2]anditya@telkomuniversity.ac.id, [3]jondri@telkomuniversity.ac.id,

Abstrak

Dengan berkembangnya sosial media terutama yang memiliki fitur untuk mengunggah foto dan gambar menyebabkan banyaknya gambar yang diunggah pada sosial media. Gambar tersebut dapat digunakan untuk membangun sistem pencarian gambar berbasis isi atau *content-base image retrieval*. Namun, banyak gambar yang diunggah tidak diberikan label atau tag sesuai dengan isi dari citra yang diunggah, sehingga sangat sulit untuk dikelola. Untuk dapat mewujudkan sistem pencarian gambar berbasis isi maka setiap obyek pada gambar harus dikenali terlebih dahulu. Jika pengenalan obyek tersebut dilakukan secara manual maka akan sangat sulit karena akan memakan waktu yang lama dan makna dari setiap orang terhadap suatu gambar berbeda, yang menimbulkan pengenalan obyek yang subyektif. Oleh karena itu dibangunlah sistem penganotasian gambar secara otomatis. Dalam penelitian ini diajukan sebuah metode *Convolutional Neural Network* untuk menangani sistem penganotasian gambar multilabel. metode *Convolutional Neural Network* telah terbukti memiliki performansi yang baik pada kasus klasifikasi gambar, ditunjukkan dengan performansi pada kasus klasifikasi gambar pada ILSVRC yang semakin membaik setiap tahunnya. Performansi tertinggi terhadap data uji pada penelitian ini adalah 81.24%.

Kata kunci : *Convolutional Neural Network*, *Image annotation*, multilabel

Abstract

With the development of social media, especially that has a feature to upload photos and images cause many of images uploaded on social media. This uploaded images could be used for developing content-base *image retrieval* system. However, many of the uploaded images are not labeled or tagged in accordance with the content of the uploaded image, so it is very difficult to manage. To be able to realize the content-based *image retrieval* system then each object in the image must be recognized first. If the annotation of the object is done manually it will be very difficult because it will take a long time and meaning of each person toward an image is different so it will cause the recognition of images is subjective. Therefore, automatic *Image annotation* system was built. In this research, we proposed a *Convolutional Neural Network* method to handle multilabel *Image annotation* system. *Convolutional Neural Network* method has been shown to have a good performance in the case of image classification, it is shown with performance in classification cases in ILSVRC which is better every years. The highest performance toward test data in this research is 81.24%.

Keywords : *Convolutional Neural Network*, *Image annotation*, multilabel

1. Pendahuluan

Dengan berkembangnya media sosial terutama media sosial yang memiliki fitur untuk mengunggah foto dan gambar seperti Instagram, Facebook, dan Flickr menyebabkan banyak gambar diunggah oleh pengguna sosial media tersebut. Dari gambar-gambar pengguna media sosial yang diunggah dapat digunakan untuk membangun sistem pencarian gambar berbasis isi atau content-base *image retrieval*. Namun, kebanyakan gambar yang diunggah tidak dianotasi atau tidak diberi keterangan yang sesuai dengan gambar oleh pengguna sehingga sangat sulit untuk dikelola. Untuk dapat mewujudkan sistem pencarian gambar berbasis isi maka setiap obyek pada gambar harus dikenali terlebih dahulu dan disimpan. Jika pengenalan obyek tersebut dilakukan secara manual maka akan sangat sulit karena akan memakan waktu yang lama dan makna dari setiap orang terhadap suatu gambar berbeda sehingga menimbulkan pengenalan obyek yang subyektif. Oleh sebab itu dikembangkanlah pengenalan dan penganotasian gambar secara otomatis yang diharapkan dapat mengenali obyek pada gambar secara obyektif berdasarkan ciri visual yang ada dalam gambar. Pada kasus Automatic *Image annotation*, banyak penelitian yang telah dilakukan hanya berfokus pada single-label saja dimana satu citra/gambar diberikan hanya 1 label seperti pada sistem anotasi gambar dengan sparse coding dan vector of locally aggregated descriptors(VLAD)[10, 3, 18, 8]. Namun, pada penerapan dalam dunia nyata, satu gambar dapat dikaitkan dengan beberapa tag atau label sehingga tidak dapat dilakukan penganotasian hanya dengan single-label saja. Pada penelitian sebelumnya, sistem penganotasian gambar multilabel telah dibangun namun masih memiliki performansi yang kurang baik. Beberapa sistem yang dibangun seperti sistem penganotasian

gambar multilabel dengan metode k-NN yang menghasilkan *overall precision* 32.29% dan *overall recall* sebesar 66.98% serta metode SVM dengan *overall precision* sebesar 22.73% dan *overall recall* sebesar 47.15%[1, 19].

Dalam penelitian ini diajukan metode *Convolutional Neural Network* untuk mengatasi permasalahan sistem penganotasian gambar yang memiliki multilabel. *Convolutional Neural Network* merupakan tipe jaringan syaraf tiruan yang menggunakan struktur tertentu seperti, convolutional layer, pooling layer, dan fully connected layer. Metode ini telah menunjukkan hasil yang menjanjikan pada permasalahan tentang klasifikasi single-label gambar[4]. Seperti pada tahun 2014 salah satu arsitektur *Convolutional Neural Network* dapat melakukan klasifikasi gambar dengan dataset bersekala besar dengan error rate sebesar 6.7% pada perlombaan ILSVRC[14]. Selain itu CNN juga sangat efektif pada pengaplikasian computer vision seperti image captioning dan object detection[11, 17].

Dalam penelitian ini diajukan metode *Convolutional Neural Network* untuk mengatasi permasalahan sistem penganotasian gambar yang memiliki multilabel. *Convolutional Neural Network* merupakan tipe jaringan syaraf tiruan yang menggunakan struktur tertentu seperti, convolutional layer, pooling layer, dan fully connected layer. Metode ini telah menunjukkan hasil yang menjanjikan pada permasalahan tentang klasifikasi single-label gambar[4]. Seperti pada tahun 2014 salah satu arsitektur *Convolutional Neural Network* dapat melakukan klasifikasi gambar dengan dataset bersekala besar dengan error rate sebesar 6.7% pada perlombaan ILSVRC[14]. Selain itu CNN juga sangat efektif pada pengaplikasian computer vision seperti image captioning dan object detection[11, 17].

Beberapa batas yang terdapat pada penelitian ini adalah :

1. dataset yang digunakan adalah dataset gambar NUS-WIDE-SCENE yang merupakan dataset gambar yang telah diberi label dengan ukuran gambar 224x224 piksel,
2. total dataset yang digunakan sejumlah 27,535 citra yang dibagi menjadi data latih sebanyak 10,913 gambar dan data uji sebanyak 16,622 citra,
3. jumlah kelas sebanyak 30 kelas.
4. Arsitektur *Convolutional Neural Network* yang digunakan adalah arsitektur VGG16.

Penelitian bertujuan untuk membangun sistem penganotasian gambar multilabel menggunakan metode *Convolutional Neural Network* dengan mengetahui parameter-parameter yang meningkatkan performa sistem. Selain itu penelitian ini akan membandingkan pengeluaran jumlah label pada sistem dan penggunaaan threshold pada penentuan label pada sistem. Performa sistem akan diukur dengan menggunakan akurasi sistem.

Urutan penulisan laporan ini adalah sebagai berikut : bagian 2 menunjukkan penelitian terkait dengan penelitian ini. Sistem yang diajukan untuk Multilabel *Image annotation* menggunakan *Convolutional Neural Network* akan dijelaskan pada bagian 3. Pada bagian 4 akan didiskusikan mengenai hasil pengujian dan evaluasi sistem. Akhirnya, kesimpulan akan dipaparkan pada bagian 5.

2. Studi Terkait

2.1 *Image annotation*

Image annotation atau anotasi gambar merupakan cabang dari *image retrieval* yang digunakan untuk memberikan label atau tag pada gambar dengan sekumpulan kata kunci berdasarkan isi dari gambar [15]. *Image annotation* menghasilkan label-label yang dapat digunakan untuk pengelompokan gambar berdasarkan isi dari gambar tersebut agar mudah dikelola. Pada penelitian sebelumnya, sistem penganotasian gambar dikembangkan menggunakan berbagai metode seperti metode ekstraksi ciri *Speed-Up Robust Features*(SURF) dengan menggunakan *classifier Support Vector Machine* dengan akurasi 91.25% pada kasus single-label, menggunakan metode klasifikasi *k-Nearest Neighbor(k-NN)* dikombinasikan dengan *Multi Non-negative Matrix Factorization* dan *Semantic Co-occurence* untuk mengingkatkan performansi dari sistem yang dibangun dengan *average precision* sebesar 40.6% dan *average recall* sebesar 42.9% pada kasus multilabel [15, 20].

2.2 Convolutional Neural Network

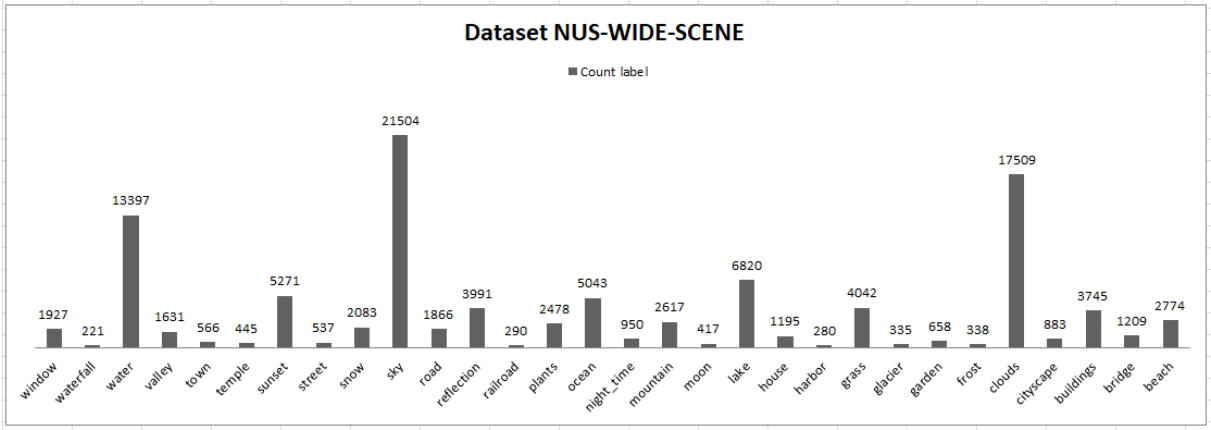
Convolutional Neural Network adalah salah satu dari metode Deep Learning. Metode ini sangat efektif digunakan pada pengaplikasian computer vision namun tidak menutup kemungkinan bahwa metode ini juga dapat digunakan untuk menyelesaikan kasus pattern recognition dan natural language processing.Desain dari *Convolutional Neural Network* terinspirasi oleh mekanisme visual pada otak. Terdapat 3 bagian dalam *Convolutional Neural Network* yaitu *convolutional layer*, *pooling layer*, dan *fully-connected layer*. Ketiga lapisan tersebut memiliki fungsi yang berbeda [4].

Beberapa penelitian mengenai *Convolutional Neural Network* telah dilakukan dan berbagai arsitektur *Convolutional Neural Network* telah diperkenalkan. Mulai dari arsitektur AlexNet yang diperkenalkan oleh Alex Krizhevsky, Ilya Sutskever, dan Geoffrey Hinton pada tahun 2012 yang digunakan untuk menyelesaikan kasus klasifikasi gambar pada kompetisi ILSVRC 2012 [9]. Pada penelitian mengenai anotasi gambar seperti pada perlombaan ImageCLEF 2014, terdapat 3 peserta yaitu MIL, MindLab, MLIA yang menggunakan ImageNet *pretrained Convolutional Neural Network* dengan beberapa modifikasi seperti pada tim MIL dan MindLab menggunakan layer keluaran dari fitur visual *Convolutional Neural Network* sedangkan pada tim MLIA,

menggunakan seluruh layer pada *pretrained Convolutional Neural Network* untuk proses pelatihan. Pada perlombaan tersebut tim MIL, MindLab, MLIA memiliki performansi tinggi namun masih belum dapat mengalahkan peringkat satu yaitu tim KDEVIR yang menggunakan *multiple* SVM *classifier* pada setiap labelnya[16].

2.3 NUS-WIDE Dataset

NUS-WIDE dataset merupakan dataset gambar yang dibuat oleh Laboratorium *Media Search* NUS. Dataset ini terdiri dari 269,648 gambar dan tag yang terkait yang diambil dari *Flickr*. Dataset NUS-WIDE dibagi kembali menjadi beberapa kategori agar dapat digunakan untuk analisis visual untuk *image annotation* dan *image retrieval* dikarenakan dataset yang terlalu besar. Salah satu dari kategori tersebut adalah NUS-WIDE-SCENE yang terdiri dari 34,926 gambar dengan tag atau label terkait[2]. Gambar 1 merupakan total perhitungan label dari Dataset NUS-WIDE-SCENE dengan 30 kelas yang akan digunakan untuk penelitian *Multilabel Image Annotation* menggunakan *Convolutional Neural Network*.

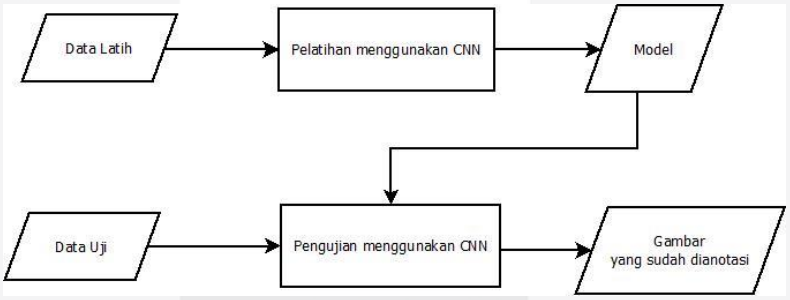


Gambar 1 Total label pada dataset NUS-WIDE-SCENE

3. Sistem yang Dibangun

3.1 Gambaran Umum Sistem

Pada penelitian ini dataset dibagi menjadi 2 yang digunakan untuk proses pelatihan sistem dan pengujian sistem. Gambar diagram 2 menunjukkan gambaran umum dari sistem, dimulai dari pelatihan model menggunakan *Convolutional Neural Network* hingga mendapatkan model dan diuji performansinya pada tahap pengujian.



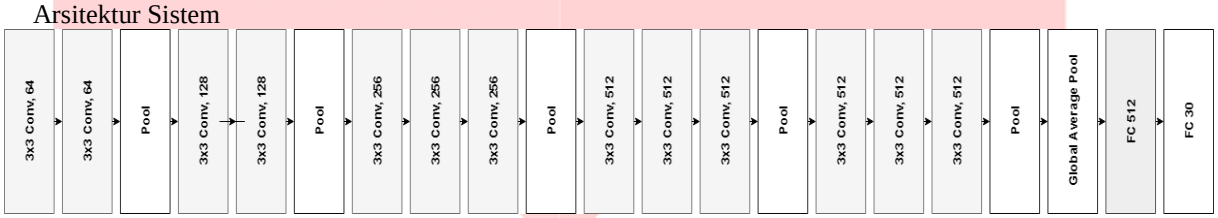
Gambar 2 Sistem yang dibangun

Dataset pelatihan NUS-WIDE-SCENE yang digunakan adalah dataset dengan label minimal 3 dengan total data pelatihan sebesar 10,913 gambar. Gambar yang memiliki label kurang dari 3 tidak akan digunakan pada penelitian ini. Selanjutnya, data latih yang telah disiapkan dilatih menggunakan jaringan pada *layer* klasifikasi (dilakukan transfer learning) selama 100 *epoch* dengan *learning rate* sebesar 0.001 dan menggunakan parameter update *adam optimizer* terhadap 30 kelas. tabel 1 merupakan contoh citra masukan dan hasil anotasi dari sistem.

Tabel 1 Dataset dan label hasil anotasi

Citra	
	
sunset, sky, clouds	*water, reflection, lake*

3.2 Arsitektur Sistem



Gambar 3 Arsitektur jaringan *Convolutional Neural Network* pada sistem yang dibangun

Arsitektur jaringan *Convolutional Neural Network* yang dibangun menggunakan arsitektur VGGNet 16 namun terdapat beberapa perubahan pada arsitektur VGG16 yang dibangun. Pada gambar 3 arsitektur VGG16 yang telah ada, layer klasifikasi pada arsitektur tersebut dihilangkan dan diganti dengan layer *Global Average Pooling* dan *fully-connected layer* dengan jumlah *neuron* 512 dan pada akhirnya ditambah dengan *fully-connected layer* dengan jumlah *neuron* sesuai jumlah kelas. Fungsi aktivasi pada *layer* terakhir diubah menjadi *Sigmoid* untuk kebutuhan klasifikasi multilabel.Pada penelitian ini digunakan *batch normalization* pada *fully-connected layer* sebelum non-linearitas yang bertujuan untuk menyamakan distribusi pada input setiap layer sehingga memungkinkan penggunaan *learning rate* yang tinggi dan mempercepat proses pelatihan[7]. Pada arsitektur sistem diterapkan juga *dropout* yang berguna untuk mencegah *overfit* pada jaringan *Convolutional Neural Network*[12].

3.3 Pelatihan menggunakan *Convolutional Neural Network*

Pada penelitian ini dilakukan 2 metode pelatihan yaitu melakukan pelatihan dari awal dan melakukan *transfer learning*.

3.3.1 Pelatihan tanpa *transfer learning*

Pada pelatihan tanpa transfer learning, arsitektur jaringan yang telah dibangun sesuai gambar 3, dilatih terhadap data latih sebanyak 10913 dengan ukuran batch sebesar 5, *learning rate* sebesar 0.01 dan dilatih menggunakan *batch normalization* dan tidak menggunakan *batch normalization* dengan parameter update yaitu *Adam optimizer* selama 60 *epoch*.

3.3.2 Pelatihan menggunakan transfer learning

Proses transfer learning ini bertujuan untuk menemukan model layer klasifikasi yang sesuai dengan dataset latih dengan menggunakan hasil learning dari model lain yang telah dilakukan pelatihan. Transfer learning dilakukan menggunakan model yang telah dilatih menggunakan data citra ImageNet yang dihilangkan layer klasifikasinya dan disambung dengan layer klasifikasi yang disesuaikan dengan penelitian ini yaitu global average pooling, fully-connected layer sebanyak 512 *neuron* dan pada layer terakhir dipasang fully-connected layer dengan *neuron* sebanyak jumlah kelas yaitu 30 *neuron*. Sistem dilatih menggunakan batch sebesar 15, dengan *dropout* 0, 0.50, dan 0.75 serta menggunakan *learning rate* 0.001 dengan Adam optimizer selama 100 *epoch*.

3.3.3 Proses Klasifikasi

Pada proses klasifikasi, dilakukan 3 cara penentuan label keluaran yaitu dengan mengeluarkan 3 label terbaik, 5 label terbaik, dan *threshold* pada masing-masing kelas. *threshold* bertujuan untuk menentukan batas penentuan label pada masing-masing *neuron output*. Penentuan *threshold* terbaik dengan cara menelusuri nilai akurasi pada setiap label dengan percobaan threshold antara 0.1 hingga 0.9 dengan penambahan 0.01 pada setiap iterasinya. Sedangkan pada penentuan label 3 dan 5 terbaik dilihat dari hasil keluaran *layer* klasifikasi dan diurutkan berdasarkan 3 atau 5 nilai keluaran *neuron* kelas terbesar.

3.3.4 Pengukuran Performansi Sistem

Dalam penelitian ini, pengukuran performansi dilakukan untuk mengevaluasi sistem yang telah dibangun. Perhitungan performansi sistem menggunakan akurasi yang dapat dituliskan sebagai berikut :

$$Accuracy = \frac{\sum_{i=1}^{m} N_i^c}{\sum_{i=1}^{m} N_i^{p \cap g}}.$$

Untuk mencegah bias pada perhitungan performansi akurasi, dilakukan pula perhitungan perclass precision yang dituliskan sebagai R^+ dan perclass precision dituliskan sebagai P^+ yang digunakan untuk melakukan perhitungan performansi per label yang dituliskan dalam persamaan berikut :

$$R^+ = \frac{1}{m}\sum_{i=1}^{m} \frac{N_i^c}{N_i^g} \qquad P^+ = \frac{1}{m}\sum_{i=1}^{m} \frac{N_i^c}{N_i^p}$$

Dilakukan juga perhitungan *overall precision*(P^A) dan *overall recall*(R^A) sebagai berikut :

$$R^A = \frac{\sum_{i=1}^{m} N_i^c}{\sum_{i=1}^{m} N_i^g} \qquad P^A = \frac{\sum_{i=1}^{m} N_i^c}{\sum_{i=1}^{m} N_i^p}$$

dimana N^c adalah jumlah prediksi label yang benar,Np merupakan jumlah label prediksi, N^g merupakan jumlah label target, dan $N^{p \wedge g}$ merupakan jumlah label yang diprediksi oleh sistem yang beririsan dengan jumlah label sebenarnya. Hasil evaluasi performansi pada setiap skenario pengujian digunakan untuk menentukan parameter terbaik pada sistem yang dibangun.

4. Evaluasi

Pada bagian ini ditampilkan hasil dari pengujian yang telah dilakukan terkait beberapa parameter yaitu batch normalization, *dropout*, dan penentuan label top-3, top-5, dan threshold serta ditampilkan pula pelatihan dengan metode transfer learning.

4.1 Perbandingan pengujian menggunakan model pelatihan menggunakan *batch normalization* dengan tidak menggunakan *batch normalization* pada *convolutional layer*

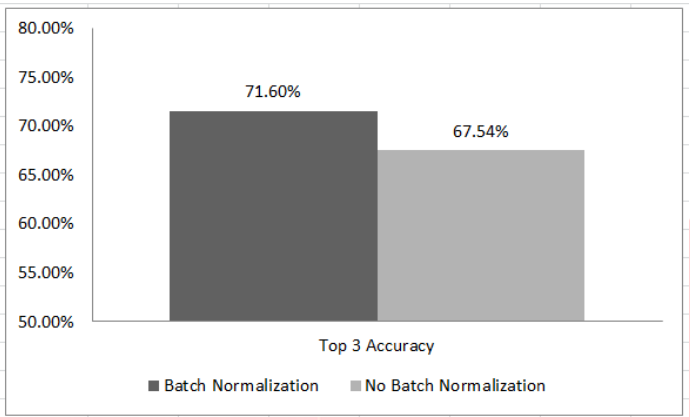
Dari arsitektur *Convolutional Neural Network* pada gambar 3, diberikan *batch normalization* pada *convolutional layer* dan pada *fully-connected layer* dan dilatih selama 60 *epoch* menggunakan *learning rate* 0.01 dengan *adam optimizer*. Berikut merupakan hasil dari pengujian terhadap arsitektur yang menggunakan *batch normalization* dan tidak menggunakan *batch normalization*. , dari tabel 2, akurasi top-3 dapat digambarkan pada grafik 4.

Tabel 2 Perbandingan *perclass precision*(P^+), *perclass recall*(R^+), *overall precision*(P^A), *overall recall*(R^A), dan akurasi top-3 pada penggunaan *batch normalization*

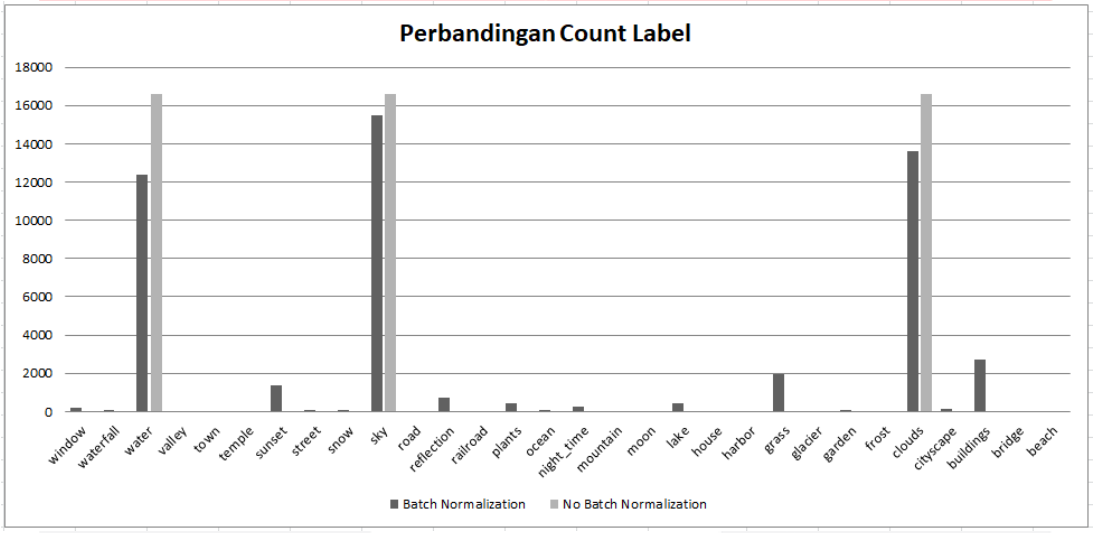
Model	P^+	R^+	P^A	R^A	Akurasi Top 3
batch normalization	23.09%	16.14%	62.01%	53.79%	71.60%
tanpa batch normalization	5.83%	10.00%	58.39%	50.64%	67.54%

Dari tabel 2 dan grafik 4 pada jaringan tanpa *batch normalization* menunjukkan akurasi top-3 sebesar 67.54% dengan total gambar yang salah dianotasi terhadap semua label target adalah 1,843 gambar. Sedangkan pada jaringan dengan *batch normalization* menunjukkan akurasi sebesar 71.60% dengan total gambar yang salah dianotasi terhadap semua label target sebesar 1,595 gambar. dari tabel 2, dapat dilihat pula bahwa nilai perclass precision, *perclass recall*, *overall precision*, dan *overall recall* pada jaringan *batch normalization* lebih tinggi dibandingkan dengan jaringan tanpa *batch normalization*. Selanjutnya, penghitungan frekuensi label pada setiap gambar dilakukan untuk memperjelas mengenai perbedaan pada penggunaan *batch normalization*. Hasil perhitungan tersebut bisa dilihat pada grafik 5.

Dari grafik 5 jika dibandingkan dengan grafik dataset 1 menunjukkan bahwa jaringan tanpa *batch normalization* hanya dapat menganotasi 3 label yaitu water, sky, clouds dan proses pembelajaran dengan 60 *epoch* masih belum dapat mengetahui ciri-ciri yang dimiliki pada label-label yang lain. Sedangkan pada jaringan yang menggunakan *batch normalization* sudah mulai terdapat label-label yang dapat dikenali oleh Convolutional Neural Network. Hal ini menunjukkan bahwa penggunaan *batch normalization* pada convolutional layer dapat mempercepat Convolutional Neural Network dalam proses pembelajaran. Pada masing-masing convolutional layer dilakukan normalisasi



Gambar 4 Perbandingan akurasi top-3 penggunaan *batch normalization*



Gambar 5 Perbandingan frekuensi label *batch normalization*

sehingga distribusi dari seluruh masukan sebelum masuk ke fungsi aktivasi akan disesuaikan. Hal inilah yang membuat perbedaan rentang pada setiap masukan ke fungsi aktivasi tidak akan berbeda jauh.

4.2 Pengujian terhadap model yang dilatih menggunakan transfer learning

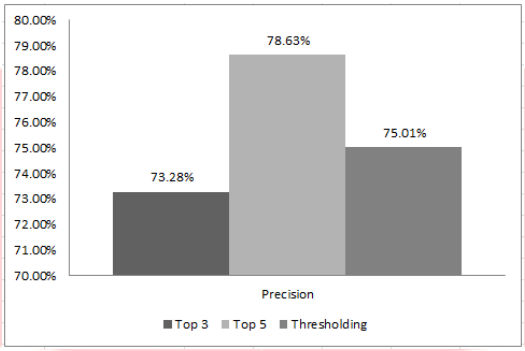
Pada pengujian dengan hasil pelatihan menggunakan transfer learning, dilakukan pengujian terhadap mekanisme pengeluaran label yaitu top-3, top-5, dan threshold, terhadap jaringan yang dilatih menggunakan *batch normalization* pada fully-connected layer dan tidak menggunakan *batch normalization*, serta penggunaan *dropout* 0, 0.5, dan 0.75 pada jaringan. Hasil pengujian terhadap mekanisme pengeluaran label top-3, top-5, dan threshold dapat dilihat pada tabel 3. Dari tabel 3, hasil akurasi dapat dibuat menjadi grafik 6. Berdasarkan grafik 6 dan tabel 3 mekanisme

Tabel 3 Perbandingan perclass precision(P^+), *perclass recall*(R^+), *overall precision*(P^A), *overall recall*(R^A) dan akurasi pada mekanisme pengeluaran label

Mekanisme Pengeluaran Label	P^+	R^+	P^A	R^A	Akurasi
Top 3	44.69%	31.68%	63.14%	54.76%	73.28%
Top 5	34.16%	50.53%	50.98%	73.68%	78.63%
Threshold	46.02%	40.96%	62.87%	67.15%	75.01%

pengeluaran label top-5 memiliki nilai akurasi sebesar 78.63% dengan *overall recall* 73.68% serta *perclass recall* 50.53% yang lebih tinggi dari top-3 yang memiliki nilai akurasi sebesar 73.28% dengan *overall recall* 54.76% serta *perclass recall* 31.68%, dan threshold sebesar 75.01% dengan *overall recall* 67.15% serta *perclass recall* 40.96%. Precision pada top-5 lebih rendah daripada mekanisme threshold dan top-3 karena rata-rata label

pada dataset adalah 3 sehingga banyak label yang dianggap salah sehingga digunakanlah perhitungan akurasi. top-5 memiliki nilai akurasi tertinggi dikarenakan pada setiap data citra yang dianotasi, sistem akan selalu mengambil 5 label dengan score terbaik, hal ini dikarenakan pada dataset NUS-WIDE-SCENE terdapat data yang memiliki



Gambar 6 Perbandingan akurasi pada mekanisme pengeluaran label

label yang tidak lengkap. Seperti pada tabel 4 dimana label target hanya 1 yaitu label *road*, sedangkan pada citra tersebut seharusnya dapat memiliki label *building*, *street*, dan *sky*.

Tabel 4 Contoh dataset yang memiliki label kurang lengkap

Citra

target : *road*
prediksi : *street, sky, road, clouds, buildings*

Selanjutnya dilakukan pengujian terhadap penggunaan *dropout* dan *batch normalization* pada fully-connected layer pada proses pelatihan transfer learning. Berikut merupakan tabel dan grafik perbandingan penggunaan probabilitas *dropout* 0, 0.5, dan 0.75 dengan *batch normalization* dan tidak menggunakan *batch normalization* terhadap akurasi top-5. Pada grafik 7 dan tabel 5 terlihat bahwa pada hasil pelatihan transfer learning, penggunaan

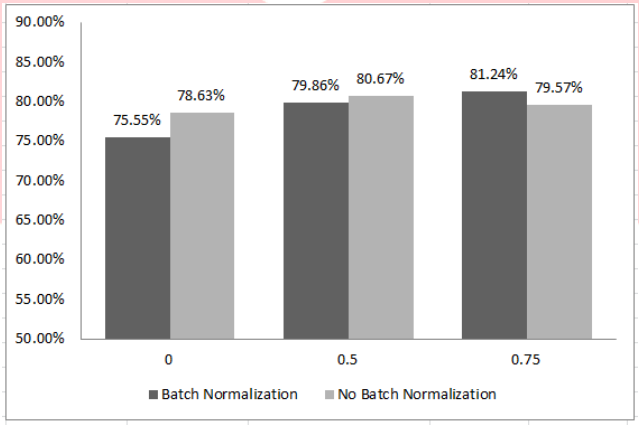
Tabel 5 Perbandingan perclass precision(P^+), *perclass recall*(R^+), *overall precision*(P^A), *overall recall*(R^A), dan akurasi pada penggunaan *batch normalization*(BN) dan *dropout(*DO)

Model	P^+	R^+	P^A	R^A	Akurasi Top 5
Batch normalization + *Dropout* 0	31.17%	47.33%	49.24%	71.17%	75.55%
Batch normalization + *Dropout* 0.5	37.78%	47.62%	51.83%	74.92%	79.86%
Batch normalization + *Dropout* 0.75	44.43%	46.60%	52.90%	76.47%	81.24%
No *Batch normalization* + *Dropout* 0	34.16%	50.53%	50.98%	73.69%	78.63%
No *Batch normalization* + *Dropout* 0.5	41.66%	46.11%	52.45%	75.82%	80.67%
No *Batch normalization* + *Dropout* 0.75	40.09%	40.09%	52.04%	75.22%	79.57%

batch normalization pada fully-connected layer memiliki akurasi lebih rendah terhadap jaringan tanpa *batch normalization* namun jika dilakukan penambahan probabilitas *dropout*, jaringan dengan *batch normalization* pada probabilitas *dropout* 0 memiliki akurasi top-5 sebesar 75.55%, pada probabilitas *dropout* 0.5 memiliki akurasi top- 5 sebesar 79.86%, dan pada probabilitas *dropout* 0.75 akurasi top-5 yang dihasilkan sebesar 81.24% dengan perclass precision, *overall precision*, dan *overall recall* yang selalu bertambah pada setiap penambahan

probabilitas *dropout*. Hal ini menunjukkan bahwa pada saat penambahan probabilitas *dropout*, performansi jaringan ini selalu naik.

Sedangkan pada jaringan tanpa *dropout* akurasi top-5 turun pada probabilitas *dropout* 0.75 walaupun penurunan tersebut tidak signifikan yaitu sebesar 1.1%, penurunan terjadi pula pada *overall precision*, *overall recall*, perclass precision, dan *perclass recall*. Hal ini terjadi karena pada saat probabilitas *dropout* 0.75 banyak neuron pada fully-connected layer akan mati sehingga pada proses pembelajaran distribusi masukan menuju fungsi aktivasi akan berubah karena pada setiap proses pembelajaran banyak neuron yang dimatikan yang membuat hasil



Gambar 7 Perbandingan akurasi top-5 pada penggunaan *dropout* dan *batch normalization*

bobot-bobot transfer learning yang masuk pada fully-connected layer mengalami perubahan distribusi. Namun dengan dilakukannya *batch normalization* pada probabilitas *dropout* yang tinggi yaitu 0.75 dapat meningkatkan performansi sistem karena *batch normalization* mengatasi perbedaan distribusi pada masing-masing neuron pada fully-connected layer.

1. Kesimpulan

Berdasarkan pengujian yang telah dilakukan dapat ditarik kesimpulan sebagai berikut :

1. pada penelitian multilabel image annotation, *Convolutional Neural Network* memiliki performansi yang baik untuk melakukan anotasi gambar multilabel dengan akurasi top-5 mencapai 81.24%,
2. mekanisme penentuan label dapat mempengaruhi performansi *Convolutional Neural Network*. Pada pelatihan menggunakan transfer learning, pengeluaran label top-3 memiliki akurasi paling kecil sebesar 73.28% sedangkan mekanisme threshold sebesar 75.01% dan top-5 sebesar 78.63%,
3. penggunaan *batch normalization* dapat meningkatkan performansi sistem dengan cara menyesuaikan distribusi pada setiap neuron masukan sebelum menuju fungsi aktivasi dan dengan dikombinasikan dengan *dropout* pada jaringan *Convolutional Neural Network* dapat meningkatkan performansi sistem hingga 81.24%.

Melihat performansi *Convolutional Neural Network* pada sistem penganotasian gambar multilabel yang cukup baik, diharapkan pada penelitian selanjutnya pengembangan sistem anotasi dengan label kelas lebih banyak dengan parameter lain seperti tebal dan ukuran filter serta arsitektur yang berbeda sangat dianjurkan untuk penelitian selanjutnya.

Daftar Pustaka

[1] S. Chengjian, S. Zhu, and Z. Shi. Image annotation via deep neural network. In Machine Vision Applications (MVA), 2015 14th IAPR International Conference on, pages 518–521. IEEE, 2015.

[2] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y.-T. Zheng. Nus-wide: A real-world web image database from national university of singapore. In Proc. of ACM Conf. on Image and Video Retrieval (CIVR'09), Santorini, Greece., July 8-10, 2009.

[3] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 248–255. IEEE, 2009.

[4] Y. Guo et al. Deep learning for visual understanding: review. 2015.

[5] K. Hara, D. Saito, and H. Shouno. Analysis of function of rectified linear unit used in deep learning. In 2015 International Joint Conference on Neural Networks (IJCNN), pages 1–8, July 2015.

[6] Y. B. Ian Goodfellow and A. Courville. Deep learning.

[7] S. Ioffe and C. Szegedy. *Batch normalization*: Accelerating deep network training by reducing internal covariate shift. CoRR, abs/1502.03167, 2015.

[8] H. J´egou, M. Douze, C. Schmid, and P. P´erez. Aggregating local descriptors into a compact image representation. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 3304–3311. IEEE, 2010.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, Advances in Neural Information Processing Systems 25, pages 1097–1105. Curran Associates, Inc., 2012.

[10] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In Computer vision and pattern recognition, 2006 IEEE computer society conference on, volume 2, pages 2169–2178. IEEE, 2006.

[11] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on, pages 512–519. IEEE, 2014.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15:1929–1958, 2014.

[13] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1):1929–1958, 2014.

[14] C. Szegedy,W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, et al. Going deeper with convolutions. Cvpr, 2015.

[15] N. M. Tuhin Shukla and S. Sharma. Automatic image annotation using surf feature. International Journal of Computer Applications (0975 - 8887), 68:4, 2013.

[16] M. Villegas and R. Paredes. Overview of the imageclef 2014 scalable concept image annotation task. In CLEF (Working Notes), pages 308–328. Citeseer, 2014.

[17] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: Lessons learned from the 2015 MSCOCO image captioning challenge. CoRR, abs/1609.06647, 2016.

[18] J.Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 3360–3367. IEEE, 2010.

[19] F.Wu et al. Weakly semi-supervised deep learning for multi-label image annotation. 1Journal of LaTex Class File, 13:9, 2014.

[20] F. Zhong and L. Ma. Image annotation using multi-view non-negative matrix factorization and semantic co-occurence. Region 10 Conference (TENCON), 2016 IEEE, 2016.