

Parallel Spatial Pyramid Convolutional Neural Network untuk Verifikasi Kekerabatan Berbasis Citra Wajah

Parallel Spatial Pyramid Convolutional Neural Network for Kinship Verification from Face Images

Reza Fuad Rachmadi^{*)}, I Ketut Eddy Purnama

Departemen Teknik Komputer, Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember
Kampus ITS Sukolilo, Surabaya, Indonesia 60111

Cara sitasi: R. F. Rachmadi and I. K. E. Purnama, "Parallel Spatial Pyramid Convolutional Neural Network untuk Verifikasi Kekerabatan Berbasis Citra Wajah," Jurnal Teknologi dan Sistem Komputer, vol. 6, no. 4, pp. 152-157, Oct. 2018. doi: 10.14710/jtsiskom.6.4.2018.152-157, [Online].

Abstract – In this paper, we proposed a parallel spatial pyramid CNN classifier for image-based kinship verification problem. Two face images that compared for kinship verification treated as input for each parallel convolutional network of our classifier. Each parallel convolutional network constructed using spatial pyramid CNN classifier. At the end of the convolutional network, we use three fully connected layers to combine each spatial pyramid CNN features and decided the final kinship prediction. We tested the proposed classifier using large-scale kinship verification dataset, called FIW dataset, consists of seven kinship problems from 1,000 families. In our approach, we treated each kinship problem as a binary classification problem with two output. We train our classifier separately for each kinship problem with same training configuration. Overall, our proposed method can achieve an average accuracy of more than 60% and outperform the baseline method.

Keywords – image-based kinship verification; parallel spatial pyramid CNN; deep spatial pyramid features

Abstrak – Pada makalah ini, klasifier paralel spatial pyramid convolutional neural network (CNN) diusulkan untuk permasalahan verifikasi kekerabatan berbasis citra wajah. Dua buah citra wajah berbeda yang digunakan untuk verifikasi kekerabatan diperlakukan sebagai input untuk dua buah jaringan paralel spatial pyramid pada klasifier yang diusulkan. Setiap jaringan paralel spatial pyramid dibangun menggunakan klasifier spatial pyramid CNN. Pada akhir jaringan konvolusi, dilakukan penggabungan seluruh fitur yang diekstrak dan dilakukan prediksi skor kekerabatan menggunakan tiga buah lapisan fully-connected. Klasifier diuji dengan menggunakan dataset verifikasi kekerabatan skala besar, yaitu dataset FIW, yang terdiri dari tujuh kategori verifikasi kekerabatan, dan dibangun dari citra wajah yang

diambil dari 1.000 keluarga yang berbeda. Klasifier dirancang dengan asumsi bahwa setiap masalah kekerabatan diperlakukan sebagai masalah klasifikasi biner dengan dua keluaran. Hasil pengujian menunjukkan bahwa klasifier yang diusulkan menghasilkan akurasi rata-rata lebih dari 60% dan mengungguli metode baseline berbasis klasifier VGG16.

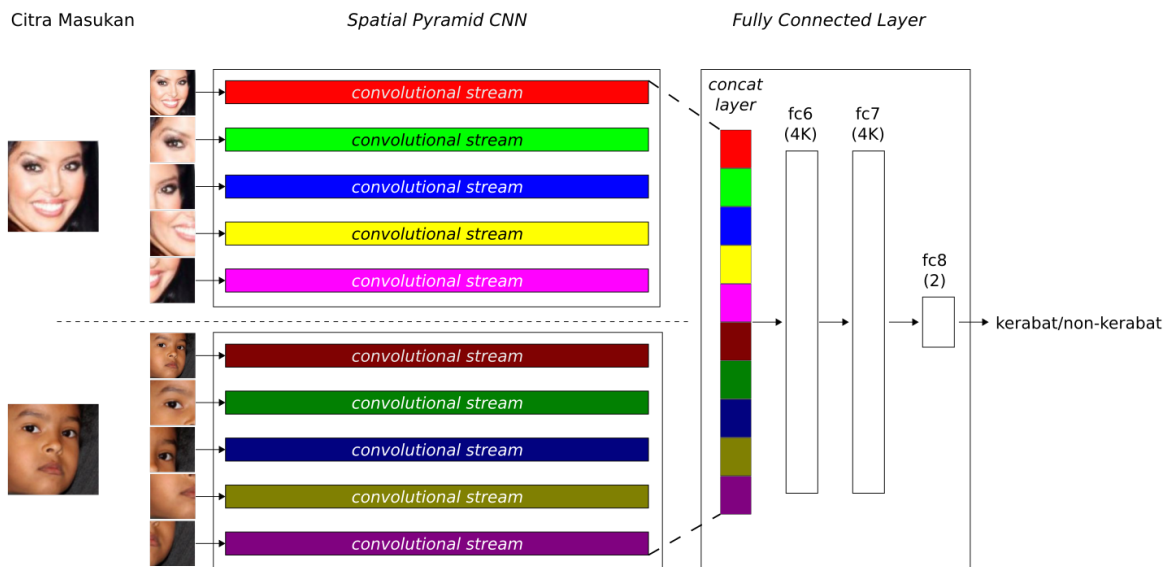
Kata Kunci – verifikasi kekerabatan berbasis citra wajah; paralel spatial pyramid CNN; fitur deep spatial pyramid

I. PENDAHULUAN

Analisa dan pemodelan citra wajah merupakan salah satu riset aktif di bidang visi komputer dan biometrika, terutama pada era internet dan *big data*. Salah satu permasalahan yang menarik adalah verifikasi kekerabatan berbasis citra wajah. Verifikasi kekerabatan berbasis citra wajah sangat berguna untuk berbagai aplikasi, di antaranya adalah penyelidikan kriminal, analisa silsilah, analisa pustaka foto, atau bahkan untuk interaksi komputer dan manusia. Verifikasi kekerabatan berbasis citra wajah akan memprediksi dua buah citra wajah orang yang berbeda memprediksi kekerabatan dari kedua citra wajah tersebut berdasarkan fitur dari citra wajah. Verifikasi kekerabatan ini mempunyai permasalahan yang sangat menantang karena ragam citra galat yang dihasilkan oleh kamera, sudut pandang wajah terhadap kamera, dan fitur wajah yang ambigu [1]-[4]. Metode yang banyak digunakan salah satunya adalah *deep learning*.

Metode *deep learning* telah digunakan untuk berbagai aplikasi, termasuk klasifikasi citra [5]-[7], deteksi objek [8]-[10], dan pengenalan suara [11], [12]. *Deep learning* menyediakan solusi menyeluruh tanpa membagi proses menjadi proses fitur ekstraksi dan proses pengenalan menggunakan metode pembelajaran mesin. Salah satu kelemahan dari *deep learning* adalah banyaknya data yang dibutuhkan untuk proses pelatihan dari nol. Strategi *fine-tuning* atau *transfer learning* dapat digunakan untuk melakukan proses pelatihan pada

^{*)}Penulis korespondensi (Reza Fuad Rachmadi)
Email: fuad@its.ac.id



Gambar 1. Diagram klasifier paralel SP-CNN yang diusulkan untuk verifikasi kekerabatan berbasis citra wajah. Citra wajah diambil dari dataset FIW [1].

data dengan skala kecil, seperti [3], [4], yang menggunakan dataset UB KinFace. Untuk permasalahan verifikasi kekerabatan berbasis citra wajah skala besar, dataset *family in the wild* (FIW) dari Robinson dkk. [1] dapat menghadirkan solusi. Dataset FIW ini berisi lebih dari 300.000 pasangan kekerabatan dengan 9 tipe kekerabatan berbeda.

Beragam teknik dikembangkan untuk menghasilkan akurasi yang tinggi dalam verifikasi kekerabatan ini, di antaranya [13] dengan *parallel spatial pyramid CNN* (SP-CNN), [2] dengan *metric learning* berbasis klasifier *denoising auto-encoder* (DAE), [3], [4] dengan *transfer learning*, dan [14] dengan *fine-tuning* klasifier. Pada makalah ini, klasifier paralel SP-CNN diusulkan untuk verifikasi kekerabatan berbasis citra wajah. Klasifier paralel SP-CNN dikembangkan berdasarkan arsitektur SP-CNN awal dalam [13]. Klasifier paralel SP-CNN diuji dengan menggunakan versi terbaru dari dataset verifikasi kekerabatan berbasis citra wajah FIW [1] yang merupakan dataset verifikasi kekerabatan berbasis citra wajah terbesar dan tersedia secara publik.

II. METODE PENELITIAN

Klasifier SP-CNN telah diperkenalkan dalam [13]. Klasifier SP-CNN terdiri dari lima jaringan dengan konfigurasi identik dengan masukan yang dipotong dari citra masukan dengan menggunakan konfigurasi *spatial pyramid*. Konfigurasi *spatial pyramid* diambil dari fitur *spatial pyramid* yang diperkenalkan oleh Lazebnik dkk. [15] yang disusun berdasarkan konfigurasi wilayah *spatial pyramid*.

Penelitian ini menggunakan klasifier paralel SP-CNN yang merupakan versi lain dari klasifier SP-CNN (Gambar 1). Paralel SP-CNN membandingkan ekstraksi fitur untuk setiap wilayah *spatial pyramid* dengan asumsi kedua wajah yang dibandingkan sejajar dan

wilayah *spatial pyramid*, meliputi seluruh wajah, mata kiri, mata kanan, pipi kiri, dan pipi kanan. Klasifier paralel SP-CNN dirancang berdasarkan arsitektur AlexNet CNN [6]. Total fitur yang diekstrak oleh paralel SP-CNN adalah 3,328 fitur untuk setiap citra wajah dengan total 6,656 fitur untuk kedua citra wajah masukan. Seluruh *neuron* pada klasifier paralel SP-CNN menggunakan fungsi aktivasi *Rectified Linear Unit* (ReLU) dalam Persamaan 1 dengan x menyatakan masukan ke sebuah neuron.

$$f(x) = \max(0, x) \quad (1)$$

Metode regularisasi *dropout* diterapkan pada lapisan *fully-connected* fc6 dan fc7 untuk mengurangi permasalahan *overfitting*. Pada bagian akhir dari klasifier, fungsi *SoftMax* digunakan untuk menormalisasi keluaran dari klasifier. Setiap kategori kekerabatan dimodelkan menjadi suatu permasalahan klasifikasi biner dengan dua keluaran, kerabat dan non-kerabat.

Untuk melakukan pelatihan dan pengujian dari klasifier paralel SP-CNN, digunakan versi terbaru dari dataset verifikasi kekerabatan berbasis citra wajah FIW [1] sebagai dataset utama serta Caffe *deep learning framework* [15] sebagai pustaka untuk melakukan pelatihan dan pengujian. Dataset FIW yang digunakan terdiri dari lebih dari 500,000 pasangan data dengan tujuh tipe kekerabatan, yaitu FS (*Father-Son/Ayah-Anak laki-laki*), FD (*Father-Daughter/Ayah-Anak perempuan*), MS (*Mother-Son/Ibu-Anak laki-laki*), MD (*Mother-Daughter/Ibu-Anak perempuan*), SS (*Sister-Sister/Antar Saudara Perempuan*), BB (*Brother-Brother/Antar Saudara laki-laki*), dan SIBS (*Siblings/Antar Saudara perempuan dan laki-laki*). Tabel 1 merupakan statistik dari dataset FIW yang digunakan untuk mengevaluasi klasifier paralel SP-CNN.

Tabel 1. Data statistik jumlah pasangan citra wajah untuk setiap tipe kekerabatan pada dataset FIW.

Tipe Kekerabatan	Jumlah Data Pelatihan	Jumlah Data Validasi	Jumlah Data Pengujian	Total Data
BB	52,483	17,342	19,946	89,771
FD	42,458	11,460	23,506	77,424
FS	53,974	13,696	45,988	113,658
MD	34,828	10,698	20,674	66,200
MS	38,312	9,816	47,954	96,082
SIBS	40,846	7,434	15,076	63,356
SS	19,286	6,218	6,524	32,028
Total	282,187	76,664	179,668	538,519

III. HASIL DAN PEMBAHASAN

Untuk memberikan variasi yang lebih banyak pada dataset yang digunakan dalam proses pelatihan, metode augmentasi data digunakan untuk memperkaya dataset FIW. Dengan menggunakan metode augmentasi data, total pasangan citra wajah meningkat 36 kali dibandingkan dengan jumlah pasangan citra wajah pada dataset FIW yang asli.

Proses augmentasi data menggunakan metode transformasi 2D *affine* dan menambahkan galat *Gaussian* untuk setiap pasangan citra wajah pada dataset FIW. Jika $\mathbf{x} \in \mathcal{R}^{3 \times 1}$ merupakan koordinat dari pixel pada citra dan $\mathbf{M} \in \mathcal{R}^{3 \times 3}$ merupakan matriks 2D transformasi *affine*, pixel masukan ditransformasi menggunakan Persamaan 2.

$$y = \mathbf{x}^T \mathbf{M} \quad (2)$$

Proses augmentasi data dilakukan dengan menggunakan sudut rotasi acak antara -10^0 sampai dengan 10^0 . Tambahan *zoom* sebesar 1,2 digunakan untuk mengurangi *pixel* hitam pada keluaran transformasi 2D *affine*. Galat *Gaussian* ditambahkan setelah proses transformasi. Proses augmentasi dilakukan sebanyak lima iterasi untuk setiap pasangan citra wajah dan menghasilkan 36 variasi pasangan citra wajah. Tabel 2 menunjukkan statistik dari data untuk proses pelatihan setelah proses augmentasi data.

Proses pelatihan klasifier paralel SP-CNN dilakukan hanya sebanyak 2 *epoch* karena data pelatihan yang sangat besar. Setiap tipe kekerabatan dilatih secara mandiri. Setiap citra wajah diubah ke resolusi 115×115 untuk proses pelatihan. Potongan acak dengan resolusi 112×112 digunakan dalam proses pelatihan. Fungsi *cross-entropy loss* (Persamaan 3) digunakan untuk proses pelatihan. Parameter θ menunjukkan bobot dari klasifier, N sebagai total contoh pasangan untuk setiap *minibatch*, p_i label dari pasangan (kerabat/non-kerabat), \hat{p}_i prediksi dari klasifier, ζ parameter *weight decay* yang digunakan untuk proses pelatihan, dan $R(\theta)$ sebagai fungsi regularisasi yang digunakan.

Tabel 2. Total pasangan citra wajah pada dataset FIW setelah proses augmentasi data.

Tipe Kekerabatan	Dataset Asli FIW	Setelah Proses Augmentasi
BB	52,483	1,889,388
FD	42,458	1,528,488
FS	53,974	1,943,064
MD	34,828	1,253,808
MS	38,312	1,379,232
SIBS	40,846	1,470,456
SS	19,286	694,296

$$E(\theta) = -\frac{1}{N} \sum_{n=1}^N p_i \log \hat{p}_i + \zeta R(\theta) \quad (3)$$

Algoritma *Stochastic Gradient Descent* (SGD) digunakan untuk proses pelatihan dengan laju pembelajaran (*learning rate*) $\alpha = 0,01$ pada permulaan proses pelatihan dan diturunkan ke $\alpha = 0,001$ setelah proses pelatihan berjalan sebanyak 1 *epoch*. Bobot dari klasifier SP-CNN diinisialisasi dengan menggunakan bobot dari klasifier AlexNet yang telah dilatih dengan menggunakan dataset ImageNet [16]. Proses pelatihan ini juga menggunakan parameter *momentum* sebesar 0,9 dan *weight decay* sebesar 0,0005 untuk mempercepat proses pelatihan. Perubahan bobot pada klasifier dihitung dengan menggunakan Persamaan 4 dan Persamaan 5. Parameter μ menyatakan parameter *momentum*, α laju pembelajaran, dan $\nabla E(\theta)$ sebagai turunan pertama dari fungsi *loss* E terhadap bobot klasifier θ . Turunan pertama dari fungsi *loss* terhadap bobot dari klasifier dihitung dengan menggunakan aturan rantai (*chain rule*) untuk perhitungan turunan yang dialirkan dari akhir dari klasifier sampai dengan layer yang akan diperbaharui.

$$V_{t+1} = \mu V_t - \alpha \nabla E(\theta) \quad (4)$$

Tabel 3. Akurasi rata-rata (%) untuk setiap tipe kekerabatan pada data validasi dan perbandingannya dengan metode *baseline*.

Metode	BB	FD	FS	MD	MS	SIBS	SS	Rata-Rata
Paralel VGG16	57.61	58.45	60.00	58.20	61.10	60.06	61.80	59.60
Paralel SP-CNN	61.53	60.58	62.01	62.84	64.70	62.13	63.62	62.49

Tabel 4. Evaluasi *precision* dan *recall* untuk setiap tipe kekerabatan dan perbandingannya dengan metode *baseline*.

Tipe Kekerabatan	Paralel SP-CNN		Paralel VGG16	
	Precision	Recall	Precision	Recall
BB	0.63334	0.50559	0.59496	0.47653
FD	0.67558	0.38377	0.68223	0.31623
FS	0.66914	0.46101	0.68434	0.37135
MD	0.71414	0.39185	0.67519	0.31595
MS	0.70687	0.47168	0.68313	0.41422
SIBS	0.64566	0.49609	0.65228	0.43099
SS	0.66035	0.52716	0.68461	0.43777
Rata-Rata	0.67215	0.46245	0.66525	0.39472

$$\theta_{t+1} = \theta_t + V_{t+1} \quad (5)$$

Untuk melakukan komparasi dengan klasifier paralel SP-CNN, pengujian juga dilakukan terhadap klasifier paralel VGG16 [17] yang digunakan sebagai metode *baseline*. Proses augmentasi data diterapkan dalam proses pelatihan paralel VGG16 untuk komparasi yang lebih adil. Bobot klasifier paralel VGG16 diinisialisasi dengan menggunakan bobot dari klasifier VGG16 yang telah dilatih dengan dataset ImageNet. Metode paralel VGG16 sebagai *baseline* mirip dengan [1], namun bobot dari klasifiernya tidak diinisialisasi dengan menggunakan bobot dari VGGFace [14].

Proses pengujian dilakukan dengan menggunakan dua buah data, yaitu data validasi dan data pengujian. Data validasi digunakan untuk memilih hasil pelatihan yang paling baik (ekivalen dengan akurasi yang tinggi) dan kemudian dianalisa dengan menggunakan data pengujian. Untuk proses pengujian, citra masukan diubah ke resolusi 124×124 dan dirangkai menjadi sepuluh potongan dengan resolusi 112×112 menggunakan konfigurasi potongan tengah, kiri atas, kiri bawah, kanan atas, kanan bawah, dan versi cermin horisontal dari potongan tersebut. Skor prediksi akhir dari klasifier paralel SP-CNN dihitung dengan mencari rata-rata skor prediksi untuk semua potongan. Evaluasi dilakukan dengan menggunakan tiga buah metrik, yaitu *precision* (Persamaan 6), *recall* (Persamaan 7), dan skor F1 (Persamaan 8). TP menyatakan kasus *true positive*, FP kasus *false positive*, dan FN kasus *false negative*.

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

Tabel 5. Evaluasi skor F1 pada data validasi dan perbandingannya dengan metode *baseline*.

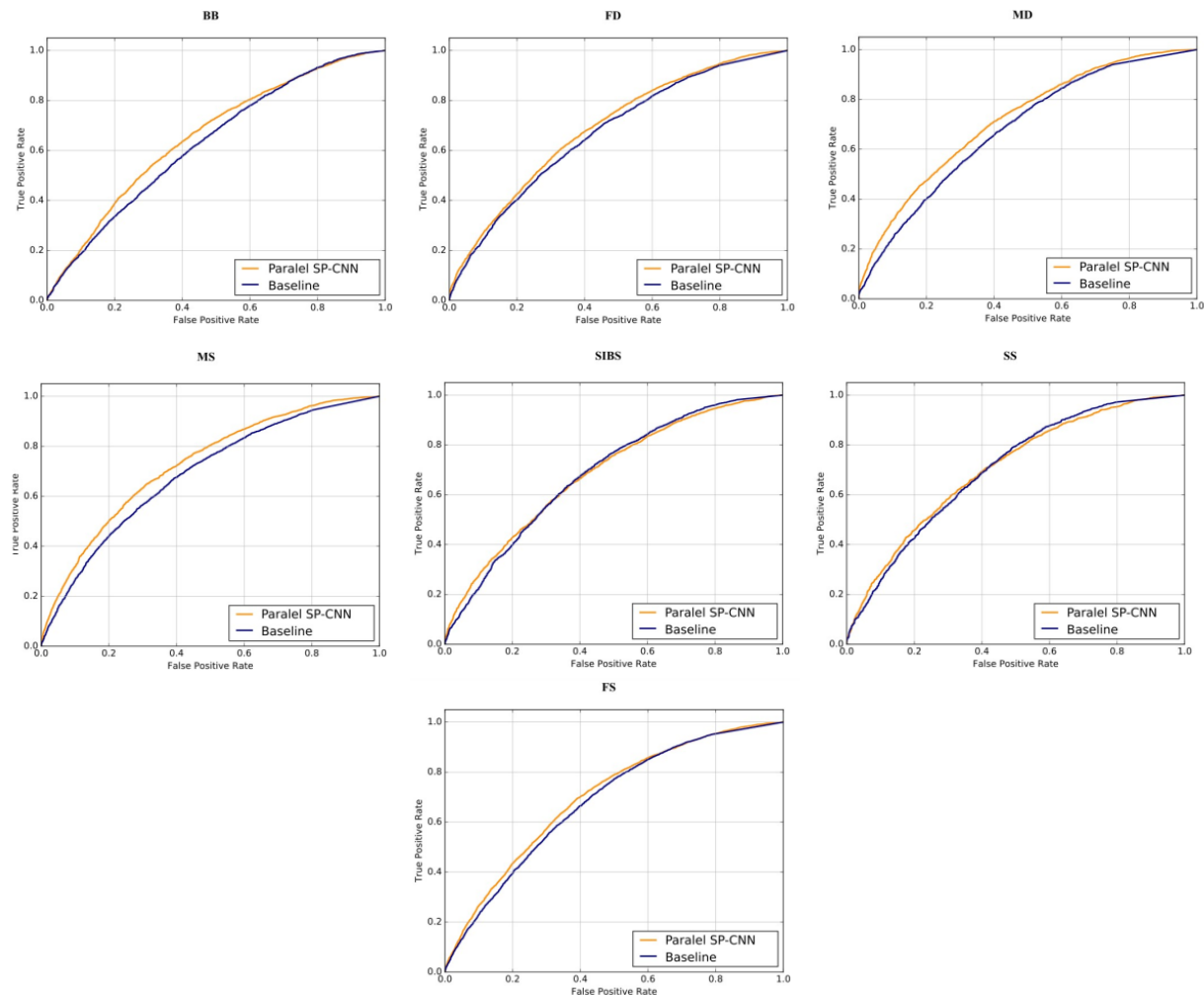
Tipe Kekerabatan	Paralel SP-CNN	Paralel VGG16
BB	0.56230	0.52920
FD	0.48948	0.43215
FS	0.54591	0.48145
MD	0.50604	0.43047
MS	0.56581	0.51573
SIBS	0.56108	0.51903
SS	0.58629	0.53405
Rata-Rata	0.54527	0.49173

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

Akurasi rata-rata yang diperoleh dari verifikasi kekerabatan menggunakan paralel SP-CNN dari dataset FIW terbaru untuk semua tipe kekerabatan adalah sebesar 62,49% masih lebih besar dari [3], [4] yang memperoleh akurasi sekitar 60% menggunakan metode transfer learning dan dataset UB KinFace dengan dataset lebih sedikit. Hasil evaluasi akurasi pada data validasi menunjukkan bahwa klasifier SP-CNN mengungguli metode *baseline* [17], yaitu menghasilkan peningkatan akurasi rata-rata sebesar ~2% (Tabel 3). Untuk semua tipe kekerabatan, metode ini juga lebih unggul daripada metode *baseline* sebesar 1-4%. Hasil ini juga menunjukkan bahwa strategi fine-tuning dari

Tabel 6. Akurasi rata-rata untuk setiap tipe kekerabatan pada data pengujian dan perbandingannya dengan metode *baseline*.

Metode	BB	FD	FS	MD	MS	SIBS	SS	Rata-Rata
Paralel VGG16	56.40	59.99	58.38	60.13	56.98	56.40	60.38	58.43
Paralel SP-CNN	60.59	61.31	61.98	62.29	57.69	61.62	63.83	61.33



Gambar 2. Grafik *precision-recall* dari klasifier paralel SP-CNN pada tujuh tipe kekerabatan pada dataset FIW.

bobot klasifier dapat menaikkan performansi verifikasi [2], [14].

Metode paralel SP-CNN ini menghasilkan *precision* yang sebanding dengan *baseline*, namun nilai *recall* klasifier paralel SP-CNN mengungguli metode *baseline* untuk semua tipe kekerabatan (Tabel 4). Analisa lebih lanjut dengan skor F1, memperlihatkan bahwa klasifier paralel SP-CNN mempunyai skor F1 yang lebih tinggi dibandingkan dengan klasifier paralel VGG16 (Tabel 5).

Hasil pengujian klasifier paralel SP-CNN pada data pengujian ditunjukkan pada Tabel 6. Akurasi rata-rata yang dihasilkan oleh klasifier paralel SP-CNN mengungguli metode *baseline* sebesar ~3%. Akurasi setiap tipe kekerabatan untuk klasifier paralel SP-CNN juga lebih tinggi dibandingkan dengan metode *baseline*. Klasifier paralel SP-CNN unggul untuk tipe kekerabatan BB, FD, MD, MS, dan FS, sedangkan untuk tipe

kekerabatan SS dan SIBS, kinerja klasifier paralel SP-CNN dengan paralel VGG16 seimbang (Gambar 2).

IV. KESIMPULAN

Klasifier paralel SP-CNN pada data FIW memiliki kurasi, *precision*, dan *recall* yang lebih baik dibandingkan dengan metode paralel VGG16 dan cukup menjanjikan untuk dikembangkan lebih lanjut sebagai alat pembantu pelacakan seseorang menggunakan *video*/citra. Arsitektur CNN mutakhir yang lain, seperti ResNet dan DenseNet, dapat digunakan dalam paralel SP-CNN untuk meningkatkan kinerja verifikasi kekerabatan berbasis citra wajah.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Dr. Supeno Mardi S. N., ST., MT. yang telah memberikan masukan dan menjadi teman diskusi penulis selama mengimplementasikan model paralel SP-CNN.

DAFTAR PUSTAKA

- [1] J. P. Robinson, M. Shao, Y. Wu, and Y. Fu, "Families in the Wild (FIW): Large-Scale Kinship Image Database and Benchmarks," In *Proceedings of the 2016 ACM on Multimedia Conference*, 2016, pp. 242–246.
- [2] S. Wang, J. P. Robinson, and Y. Fu, "Kinship Verification on Families in the Wild with Marginalized Denoising Metric Learning," In *12th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 2017, pp. 216–221.
- [3] S. Xia, M. Shao, and Y. Fu, "Kinship verification through transfer learning," In *International Joint Conference on Artificial Intelligent (IJCAI)*, 2011, pp. 2539–2544.
- [4] S. Xia, M. Shao, J. Luo, and Y. Fu, "Understanding Kin Relationships in a Photo," *IEEE Transactions on Multimedia*, Vol. 14, No. 4, pp. 1046–1056, 2012.
- [5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016. pp. 770–778.
- [6] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," In *Advances in Neural Information Processing Systems (NIPS)*, 2012. pp. 1097–1105.
- [7] R. F. Rachmadi, Y. Komokata, K. Uchimura, and G. Koutaki, "Road sign classification system using cascade convolutional neural network," *International Journal of Innovative Computing, Information and Control (IJICIC)*, Vol. 13, No.1, pp. 95–109, 2017.
- [8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," In *European conference on computer vision (ECCV)*, 2016. pp. 21–37.
- [9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. pp. 779–788.
- [10] J. Redmon, and A. Farhadi, "YOLO9000: Better, Faster, Stronger," In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 6517–6525.
- [11] Y. Qian and P. C. Woodland, "Very deep convolutional neural networks for robust speech recognition," In *Spoken Language Technology Workshop (SLT)*, 2016. pp. 481–488.
- [12] Y. Zhang, M. Pezeshki, P. Brakel, S. Zhang, C. Laurent, Y. Bengio, and A. Courville, "Towards End-to-End Speech Recognition with Deep Convolutional Neural Networks," In *Proceedings of Interspeech*, 2016. pp. 410–414.
- [13] R. F. Rachmadi, K. Uchimura, and G. Koutaki, "Spatial Pyramid Convolutional Neural Network for Social Event Detection in Static Image. In *11th International Student Conference on Advanced Science and Technology (ICAST)*, 2016.
- [14] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," In *Proceedings of the British Machine Vision Conference (BMVC)*, 2015.
- [15] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional Architecture for Fast Feature Embedding," In *Proceedings of the 22nd ACM International Conference on Multimedia (MM '14)*, 2014. pp. 675–678.
- [16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, Vol. 115, No. 3, pp. 211–252, 2015.
- [17] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014.