

**DETEKSI FOKUS ATENSI VISUAL PESERTA DIDIK DI RUANG
KELAS BERDASARKAN POSE KEPALA MENGGUNAKAN
EFFICIENTNETV2 DENGAN *SEAT POSITION EMBEDDING***



SKRIPSI

diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar sarjana
Komputer Program Studi Ilmu Komputer

Oleh:

Ananda Myzza Marhelio

2100702

**PROGRAM STUDI ILMU KOMPUTER
FAKULTAS PENDIDIKAN MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS PENDIDIKAN INDONESIA
BANDUNG
2025**

**DETEKSI FOKUS ATENSI VISUAL PESERTA DIDIK DI RUANG
KELAS BERDASARKAN POSE KEPALA MENGGUNAKAN
EFFICIENTNETV2 DENGAN *SEAT POSITION EMBEDDING***

Oleh
Ananda Myzza Marhelio
2100702

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar
Sarjana pada Fakultas Pendidikan Matematika dan Ilmu Pengetahuan Alam

© Ananda Myzza Marhelio
Universitas Pendidikan Indonesia
Juli 2025

Hak Cipta dilindungi undang-undang.
Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian,
dengan dicetak ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis.

ANANDA MYZZA MARHELIO

DETEKSI FOKUS ATENSI VISUAL PESERTA DIDIK DI RUANG KELAS
BERDASARKAN POSE KEPALA MENGGUNAKAN EFFICIENTNETV2
DENGAN *SEAT POSITION EMBEDDING*

Disetujui dan disahkan oleh pembimbing:

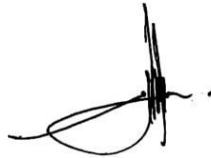
Pembimbing I



Prof. Dr. Munir, M.IT.

NIP. 196603252001121001

Pembimbing II



Yaya Wihardi, S. Kom., M. Kom.

NIP. 198903252015041001

Mengetahui,

Ketua Program Studi Ilmu Komputer



Dr. Muhamad Nursalman, M.T.

NIP. 197909292006041002

DETEKSI FOKUS ATENSI VISUAL PESERTA DIDIK DI RUANG KELAS
BERDASARKAN POSE KEPALA MENGGUNAKAN EFFICIENTNETV2
DENGAN *SEAT POSITION EMBEDDING*

Oleh
Ananda Myzza Marhelio
2100702

ABSTRAK

Pemahaman terhadap arah perhatian visual siswa sangat penting dalam mengevaluasi keterlibatan mereka selama pembelajaran di kelas. *Head Pose Estimation* (HPE) menjadi metode yang efektif dalam mengidentifikasi focus atensi, namun penerapannya di kelas nyata sering terkendala oleh kualitas citra rendah dan posisi duduk siswa yang bervariasi yang menyebabkan metode regresi untuk memprediksi *landmark* wajah atau sudut Euler tidak optimal. Penelitian ini menggunakan pendekatan klasifikasi berbasis citra sebagai alternatif dan mengusulkan modifikasi arsitektur EfficientNetV2-S dengan menambahkan *Seat Position Embedding* (SPE) sebagai konteks spasial untuk meningkatkan akurasi klasifikasi pose kepala siswa. Set data dikembangkan melalui rekaman langsung di kelas dan diproses menjadi 4.574 gambar pose kepala dengan lima label arah (atas, bawah, depan, kanan, kiri). Evaluasi dilakukan pada beberapa arsitektur CNN dengan dan tanpa SPE. Hasil menunjukkan bahwa penambahan SPE pada model yang diusulkan memperoleh akurasi sebesar 83,25%, melebihi akurasi model *baseline* pada 82,53%. Pendekatan ini terbukti efisien dalam mengurangi ambiguitas visual dan memberikan interpretasi lebih akurat terhadap atensi siswa.

Kata kunci: Deteksi Atensi Siswa, EfficientNetV2, Estimasi Pose Kepala, Ruang Kelas, Siswa

VISUAL ATTENTION FOCUS DETECTION OF STUDENTS IN
CLASSROOM BASED ON HEAD POSE USING EFFICIENTNETV2 WITH
SEAT POSITION EMBEDDING

Arranged by
Ananda Myzza Marhelio
2100702

ABSTRACT

Understanding students' visual attention direction is essential for evaluating their engagement during classroom learning. Head Pose Estimation (HPE) is an effective method for identifying attention focus, however, its application in real classroom settings is often hindered by low image quality and varied student seating positions, which makes regression-based methods for predicting facial landmarks or Euler angles suboptimal. This study adopts an image-based classification approach as an alternative and proposes a modification of the EfficientNetV2-S architecture by integrating Seat Position Embedding (SPE) as spatial context to improve the accuracy of head pose classification. The dataset was developed from direct classroom recordings and processed into 4,574 head pose images with five directional labels (up, down, front, right, left). Several CNN architectures were evaluated with and without SPE. The results show that the proposed model with SPE achieved an accuracy of 83.25%, surpassing the baseline model's accuracy of 82.53%. This approach has proven effective in reducing visual ambiguity and providing a more accurate interpretation of students' attention.

Keywords: Classroom, EfficientNetV2, Head Pose Estimation, Student, Student Attention Detection

DAFTAR ISI

PERNYATAAN BEBAS PLAGIARISME	iii
KATA PENGANTAR.....	iv
UCAPAN TERIMA KASIH.....	v
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL.....	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	5
1.5 Batasan Penelitian	5
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Peta Literatur.....	7
2.2 <i>Head Pose Estimation</i>	7
2.2.1 Regresi.....	8
2.2.2 Klasifikasi	10
2.3 <i>Computer Vision</i>	13
2.3.1 Klasifikasi Gambar.....	15
2.3.2 Deteksi Objek.....	16
2.4 <i>Natural Language Proccessing</i>	17
2.4.1 <i>Contextual Embeddings</i>	18
2.5 <i>Deep Learning</i>	19
2.5.1 <i>Transfer Learning</i>	21
2.5.2 EfficientNetV2	22
2.5.3 ConvNeXt	23
2.5.4 MobileNetV3.....	25
2.5.5 ResNet50.....	27
2.6 <i>Class Weight</i>	29
2.7 Metrik Evaluasi	30

2.7.1	Metrik Evaluasi Klasifikasi.....	30
2.7.2	<i>Average Precision</i>	31
2.7.3	AUC-ROC.....	32
2.8	Penelitian Terkait	33
BAB III METODE PENELITIAN		38
3.1	Desain Penelitian.....	38
3.1.1	Perumusan Masalah	38
3.1.2	Tinjauan Pustaka	39
3.1.3	Pengumpulan Set Data	39
3.1.4	Praproses Data.....	40
3.1.4.1	Set Data Klasifikasi.....	40
3.1.4.2	Augmentasi Data.....	41
3.1.4.3	Set Data Deteksi.....	42
3.1.5	Pengembangan Model.....	43
3.1.5.1	Arsitektur EfficientNetV2.....	43
3.1.5.2	Arsitektur CNN Lainnya.....	45
3.1.5.3	Eksperimen Konfigurasi Model	46
3.1.6	Evaluasi Model.....	47
3.1.7	Analisis dan Kesimpulan.....	47
3.2	Kebutuhan Perangkat	48
BAB IV HASIL DAN PEMBAHASAN.....		49
4.1	Hasil	49
4.1.1	Set Data	49
4.1.1.1	Set Data Klasifikasi.....	50
4.1.1.2	Set Data Deteksi.....	50
4.1.2	Praproses Data.....	51
4.1.2.1	Praproses Set Data Klasifikasi	51
4.1.2.2	Augmentasi Data.....	54
4.1.2.3	Praproses Set Data Deteksi	55
4.1.3	Klasifikasi Pose Kepala.....	58
4.1.3.1	Eksperimen Augmentasi	60
4.1.3.2	Eksperimen <i>Transfer Learning</i>	64
4.1.3.3	Eksperimen Tipe Model EfficientNetV2	68
4.1.3.4	Eksperimen <i>Class Weight</i>	71

4.1.3.5	EfficientNetV2 dengan <i>Seat Position Embedding</i>	74
4.1.3.6	Model Berbasis CNN dengan <i>Seat Position Embedding</i>	80
4.1.4	Deteksi Pose Kepala.....	82
4.2	Pembahasan.....	87
4.2.1	Klasifikasi Pose Kepala.....	87
4.2.1.1	Eksperimen Augmentasi	88
4.2.1.2	Eksperimen <i>Transfer Learning</i>	91
4.2.1.3	Eksperimen Tipe Model EfficientNetV2	93
4.2.1.4	Eksperimen <i>Class Weight</i>	95
4.2.1.5	EfficientNetV2 dengan <i>Seat Position Embedding</i>	97
4.2.1.6	Model Berbasis CNN dengan <i>Seat Position Embedding</i>	100
4.2.2	Deteksi Pose Kepala.....	103
BAB V SIMPULAN DAN SARAN		109
5.1	Simpulan	109
5.2	Saran.....	110
DAFTAR PUSTAKA		112

DAFTAR GAMBAR

Gambar 2.1 Peta Literatur	7
Gambar 2.2 Tiga <i>Degrees-of-Freedom</i> pada Kepala Manusia	8
Gambar 2.3 <i>Head Pose Estimation</i> dengan Pendekatan <i>Keypoints-Based</i>	9
Gambar 2.4 <i>Head Pose Estimation</i> dengan Pendekatan <i>Non Keypoints-Based</i>	10
Gambar 2.5 <i>Head Pose Estimation</i> dengan Pendekatan Klasifikasi Berdasarkan Sudut	11
Gambar 2.6 <i>Head Pose Estimation</i> dengan Pendekatan Klasifikasi Berdasarkan Perilaku Kepala: (a) <i>gazing</i> , (b) <i>non_gazing</i>	12
Gambar 2.7 Alur Kerja Traditional <i>Computer Vision</i> (a) dan <i>Deep Learning</i> (b).	15
Gambar 2.8 Ilustrasi Model Arsitektur <i>Deep Learning</i>	20
Gambar 2.9 Ilustrasi Model Arsitektur EfficientNetV2.....	22
Gambar 2.10 Ilustrasi Model Arsitektur ConvNeXt	24
Gambar 2.11 Ilustrasi Model Arsitektur MobileNetV3	26
Gambar 2.12 Arsitektur VGG-19 (Kiri), <i>Plain Network</i> (Tengah), <i>Residual Network</i> (Kanan)	28
Gambar 3.1 Desain Penelitian.....	38
Gambar 3.2 Arsitektur EfficientNetV2 dengan <i>Seat Position Embedding</i> (SPE) .	43
Gambar 3.3 Arsitektur Berbasis CNN dengan <i>Seat Position Embedding</i> (SPE)...	45
Gambar 4.1 Potongan Gambar dari Ekstraksi Video	50
Gambar 4.2 Contoh Kesalahan Model Deteksi Kepala	52
Gambar 4.3 Sampel Gambar dari Lima Kelas Label dalam Set Data Arah Pose Kepala: (a) atas, (b) bawah, (c) depan, (d) kanan, (e) kiri	53
Gambar 4.4 Distribusi Set Data Pelatihan dan Validasi	54
Gambar 4.5 Hasil Augmentasi Data.....	55
Gambar 4.6 Format JSON untuk Hasil Anotasi COCO.....	56
Gambar 4.7 Distribusi Kelas Arah Pose Kepala pada Set Data Deteksi.....	58
Gambar 4.8 Format JSON untuk Hasil Anotasi COCO.....	58
Gambar 4.9 Desain Eksperimen.....	59
Gambar 4.10 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-S tanpa Augmentasi Data	60

Gambar 4.11 <i>Confusion Matrix</i> untuk EfficientNetV2-S tanpa Augmentasi Data	61
Gambar 4.12 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-S dengan Augmentasi Data	62
Gambar 4.13 <i>Confusion Matrix</i> untuk EfficientNetV2-S dengan Augmentasi Data	63
Gambar 4.14 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-S <i>From Scratch</i>	65
Gambar 4.15 <i>Confusion Matrix</i> untuk EfficientNetV2-S <i>From Scratch</i>	65
Gambar 4.16 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-S dengan <i>Transfer Learning</i>	66
Gambar 4.17 <i>Confusion Matrix</i> untuk EfficientNetV2-S dengan <i>Transfer Learning</i>	67
Gambar 4.18 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-Medium	69
Gambar 4.19 <i>Confusion Matrix</i> untuk EfficientNetV2-Medium	69
Gambar 4.20 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-Small	70
Gambar 4.21 <i>Confusion Matrix</i> untuk EfficientNetV2-Small	71
Gambar 4.22 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-S Tanpa <i>Class Weight</i>	72
Gambar 4.23 <i>Confusion Matrix</i> untuk EfficientNetV2-S Tanpa <i>Class Weight</i>	72
Gambar 4.24 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-S Dengan <i>Class Weight</i>	73
Gambar 4.25 <i>Confusion Matrix</i> untuk EfficientNetV2-S Dengan <i>Class Weight</i> ...	74
Gambar 4.26 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-S Tanpa SPE	76
Gambar 4.27 <i>Confusion Matrix</i> untuk EfficientNetV2-S Tanpa SPE	76
Gambar 4.28 Grafik Akurasi dan <i>Loss</i> untuk EfficientNetV2-S Dengan SPE	77
Gambar 4.29 <i>Confusion Matrix</i> untuk EfficientNetV2-S Dengan SPE	77
Gambar 4.30 Gambar Grafik AUC-ROC EfficientNetV2-S: (a) Dengan SPE, (b) Tanpa SPE	78
Gambar 4.31 Perbandingan Hasil Prediksi Benar dan Salah oleh EfficientNetV2-S dengan SPE	79
Gambar 4.32 Gambar <i>Confusion Matrix</i> ResNet50: (a) Dengan SPE, (b) Tanpa SPE	81

Gambar 4.33 Hasil Deteksi Pose Kepala pada Set Data Deteksi	85
Gambar 4.34 <i>Output</i> Video	86
Gambar 4.35 <i>Output</i> Berupa CSV dan Grafik Deret Waktu.....	87

DAFTAR TABEL

Tabel 4.1 Distribusi Set Data	54
Tabel 4.2 Deskripsi Atribut Anotasi COCO	56
Tabel 4.3 Perbandingan Performa Model Dengan dan Tanpa Augmentasi Data ..	60
Tabel 4.4 Perbandingan Performa Model <i>From Scratch</i> dan <i>Transfer Learning</i> ..	64
Tabel 4.5 Perbandingan Performa Tipe Model EfficientNetV2	68
Tabel 4.6 Perbandingan Performa Model Dengan dan Tanpa <i>Class Weight</i>	71
Tabel 4.7 Perbandingan Performa EfficientNetV2-S Dengan dan Tanpa SPE	74
Tabel 4.8 Perbandingan Parameter EfficientNetV2-S Dengan dan Tanpa SPE	75
Tabel 4.9 Perbandingan Performa Model Berbasis CNN Dengan dan Tanpa SPE	80
Tabel 4.10 Perbandingan Parameter dan Waktu Pelatihan Model Berbasis CNN	82
Tabel 4.11 Perbandingan Performa Model Deteksi Kepala pada Set Data Deteksi	83
Tabel 4.12 Perbandingan Performa Model Deteksi Pose Kepala	84
Tabel 4.13 Perbandingan Kecepatan Inferensi Model Deteksi Arah Pose Kepala	85

DAFTAR PUSTAKA

- Abd Elaziz, M., Dahou, A., Alsaleh, N. A., Elsheikh, A. H., Saba, A. I., & Ahmadein, M. (2021). Boosting COVID-19 Image Classification Using MobileNetV3 and Aquila Optimizer Algorithm. *Entropy*, 23(11), 1383. <https://doi.org/10.3390/e23111383>
- Aldakhil, L. A., & Almutairi, A. A. (2024). Multi-Fruit Classification and Grading Using a Same-Domain Transfer Learning Approach. *IEEE Access*, 12, 44960–44971. <https://doi.org/10.1109/ACCESS.2024.3379276>
- Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep Reinforcement Learning: A Brief Survey. *IEEE Signal Processing Magazine*, 34(6), 26–38. <https://doi.org/10.1109/MSP.2017.2743240>
- Bakirarar, B., & Elhan, A. H. (2023). Class Weighting Technique to Deal with Imbalanced Class Problem in Machine Learning: Methodological Research. *Turkiye Klinikleri Journal of Biostatistics*, 15(1), 19–29. <https://doi.org/10.5336/biostatic.2022-93961>
- Ballard, D. H., & Brown, C. M. (1982). *Computer vision*. Prentice Hall Professional Technical Reference.
- Brownlee, J. (2019). *Deep learning for computer vision: image classification, object detection, and face recognition in python*. Machine Learning Mastery.
- Canedo, D., Trifan, A., & Neves, A. J. R. (2018). Monitoring Students' Attention in a Classroom Through Computer Vision. In J. Bajo, J. M. Corchado, E. M. Navarro Martínez, E. Osaba Icedo, P. Mathieu, P. Hoffa-D'kabrowska, E. del Val, S. Giroux, A. J. M. Castro, N. Sánchez-Pi, V. Julián, R. A. Silveira, A. Fernández, R. Unland, & R. Fuentes-Fernández (Eds.), *Highlights of Practical Applications of Agents, Multi-Agent Systems, and Complexity: The PAAMS Collection* (pp. 371–378). Springer International Publishing. https://doi.org/10.1007/978-3-319-94779-2_32
- Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student Engagement and Student Learning: Testing the Linkages*. *Research in Higher Education*, 47(1), 1–32. <https://doi.org/10.1007/s11162-005-8150-9>
- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-End Object Detection with Transformers. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 12346 LNCS, 213–229. https://doi.org/10.1007/978-3-030-58452-8_13
- Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 1, 886–893. <https://doi.org/10.1109/CVPR.2005.177>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings*

- of the 2019 Conference of the North, 1, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Dhingra, N. (2021). HeadPosr: End-to-end Trainable Head Pose Estimation using Transformer Encoders. *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*, 1–8. <https://doi.org/10.1109/FG52635.2021.9667080>
- Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, 88(2), 303–338. <https://doi.org/10.1007/s11263-009-0275-4>
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
- Hagenauer, G., Hascher, T., & Volet, S. E. (2015). Teacher emotions in the classroom: associations with students' engagement, classroom discipline and the interpersonal teacher-student relationship. *European Journal of Psychology of Education*, 30(4), 385–403. <https://doi.org/10.1007/s10212-015-0250-0>
- Hambali, Y. A., Megasari, R., & Santoso, R. R. (2023). Implementasi Metode Machine Learning menggunakan Algoritma Evolving Artificial Neural Network pada Kasus Prediksi Diagnosis Diabetes. *Jurnal Aplikasi Dan Teori Ilmu Komputer*, 6(1), 9–20. <https://doi.org/10.17509/jatikom.v6i1.56536>
- Harris, C., & Stephens, M. (1988). A Combined Corner and Edge Detector. *Proceedings of the Alvey Vision Conference*, 23.1-23.6.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Heaton, J. (2018). Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning. *Genetic Programming and Evolvable Machines*, 19(1–2), 305–307. <https://doi.org/10.1007/s10710-017-9314-z>
- Howard, A., Sandler, M., Chu, G., Chen, L.-C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V., & Adam, H. (2019). Searching for MobileNetV3. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 1314–1324. <https://doi.org/10.1109/ICCV.2019.00140>
- Jiang, B., Xu, W., Guo, C., Liu, W., & Cheng, W. (2019). A classroom concentration model based on computer vision. *Proceedings of the ACM Turing Celebration Conference - China*, 1–6. <https://doi.org/10.1145/3321408.3322856>
- Kang, C., Jin, S., Zhong, Z., Li, K., & Zeng, X. (2024). Predicting classroom activity index through multi-scale head posture classification network. *Journal of Intelligent & Fuzzy Systems*, 46(4), 8169–8183. <https://doi.org/10.3233/JIFS-237970>

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.
- Lafuente, D., Cohen, B., Fiorini, G., García, A. A., Bringas, M., Morzan, E., & Onna, D. (2021). A Gentle Introduction to Machine Learning for Chemists: An Undergraduate Workshop Using Python Notebooks for Visualization, Data Processing, Analysis, and Modeling. *Journal of Chemical Education*, 98(9), 2892–2898. <https://doi.org/10.1021/acs.jchemed.1c00142>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision -- ECCV 2014* (pp. 740–755). Springer International Publishing. https://doi.org/10.1007/978-3-319-10602-1_48
- Liu, H., Li, D., Wang, X., Liu, L., Zhang, Z., & Subramanian, S. (2021). Precise head pose estimation on HPD5A database for attention recognition based on convolutional neural network in human-computer interaction. *Infrared Physics & Technology*, 116, 103740. <https://doi.org/10.1016/j.infrared.2021.103740>
- Liu, Q., Kusner, M. J., & Blunsom, P. (2020). *A Survey on Contextual Embeddings*. <http://arxiv.org/abs/2003.07278>
- Liu, T., Wang, J., Yang, B., & Wang, X. (2021). NGDNet: Nonuniform Gaussian-label distribution learning for infrared head pose estimation and on-task behavior understanding in the classroom. *Neurocomputing*, 436, 210–220. <https://doi.org/10.1016/j.neucom.2020.12.090>
- Liu, T., Yang, B., Liu, H., Ju, J., Tang, J., Subramanian, S., & Zhang, Z. (2022). GMDL: Toward precise head pose estimation via Gaussian mixed distribution learning for students' attention understanding. *Infrared Physics & Technology*, 122, 104099. <https://doi.org/10.1016/j.infrared.2022.104099>
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. In *Computer Vision – ECCV 2016* (pp. 21–37). Springer International Publishing. https://doi.org/10.1007/978-3-319-46448-0_2
- Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A ConvNet for the 2020s. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2022-June*, 11966–11976. <https://doi.org/10.1109/CVPR52688.2022.01167>
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 2, 1150–1157 vol.2. <https://doi.org/10.1109/ICCV.1999.790410>

- Murphy-Chutorian, E., & Trivedi, M. M. (2009). Head Pose Estimation in Computer Vision: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(4), 607–626. <https://doi.org/10.1109/TPAMI.2008.106>
- O' Mahony, N., Murphy, T., Panduru, K., Riordan, D., & Walsh, J. (2016). Adaptive process control and sensor fusion for process analytical technology. *2016 27th Irish Signals and Systems Conference (ISSC)*, 1–6. <https://doi.org/10.1109/ISSC.2016.7528449>
- O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., Riordan, D., & Walsh, J. (2020). *Deep Learning vs. Traditional Computer Vision* (pp. 128–144). https://doi.org/10.1007/978-3-030-17795-9_10
- Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2014). Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks. *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 1717–1724. <https://doi.org/10.1109/CVPR.2014.222>
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10), 1872–1897. <https://doi.org/10.1007/s11431-020-1647-3>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016-Decem, 779–788. <https://doi.org/10.1109/CVPR.2016.91>
- Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- Ruiz, N., Chong, E., & Rehg, J. M. (2018). Fine-Grained Head Pose Estimation Without Keypoints. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2155–215509. <https://doi.org/10.1109/CVPRW.2018.00281>
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
- Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLOS ONE*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Setiawan, W., Nursalman, M., Munir, & Anugrah, R. D. (2017). Determine focus based on eye gazing direction. *2017 3rd International Conference on Science*

- in *Information Technology (ICSITech)*, 577–581.
<https://doi.org/10.1109/ICSITech.2017.8257179>
- Shao, S., Zhao, Z., Li, B., Xiao, T., Yu, G., Zhang, X., & Sun, J. (2018). *CrowdHuman: A Benchmark for Detecting Human in a Crowd*.
<http://arxiv.org/abs/1805.00123>
- Shen, J., Qin, X., & Zhou, Z. (2022). Head Pose Estimation In Classroom Scenes. *2022 4th International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*, 343–349.
<https://doi.org/10.1109/AIAM57466.2022.00072>
- Simonyan, K., & Zisserman, A. (2015). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. <http://arxiv.org/abs/1409.1556>
- Smith, B., & Dyer, C. (2016). Pose-Robust 3D Facial Landmark Estimation from a Single 2D Image. In R. C. Wilson, E. R. Hancock, & W. A. P. Smith (Eds.), *Proceedings of the British Machine Vision Conference 2016* (pp. 18.1-18.12). British Machine Vision Association. <https://doi.org/10.5244/C.30.18>
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Stiefelhausen, R., & Zhu, J. (2002). Head orientation and gaze direction in meetings. *CHI '02 Extended Abstracts on Human Factors in Computing Systems*, 858–859. <https://doi.org/10.1145/506443.506634>
- Szeliski, R. (2022). *Computer vision: algorithms and applications*. Springer Nature.
- Tan, M., & Le, Q. V. (2021). EfficientNetV2: Smaller Models and Faster Training. *Proceedings of Machine Learning Research*, 139, 10096–10106. <https://doi.org/10.48550/arXiv.2104.00298>
- Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192. <https://doi.org/10.1016/j.aci.2018.08.003>
- Uçar, M. U., & Özdemir, E. (2022). Recognizing Students and Detecting Student Engagement with Real-Time Image Processing. *Electronics*, 11(9), 1500. <https://doi.org/10.3390/electronics11091500>
- Umbaugh, S. E. (2010). *Digital image processing and analysis: human and computer vision applications with CVIPtools*. CRC press.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All you Need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

- Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognition Letters*, 141, 61–67. <https://doi.org/10.1016/j.patrec.2020.07.042>
- Wang, S., & Su, Z. (2019). *Metamorphic Testing for Object Detection Systems*. <http://arxiv.org/abs/1912.12162>
- Wihardi, Y., Junaeti, E., Setiawan, W., Wahyudin, W., & Erlangga, E. (2022). Smart Classroom System (SCS) Berbasis Kamera Untuk Memantau Keadaan Peserta Didik. *INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS: Journal of Information System*, 6(1), 67. <https://doi.org/10.51211/isbi.v6i1.1771>
- Wikarsa, A., Sukamto, R. A., & Wihardi, Y. (2020). Estimasi Pose Kepala Menggunakan Histogram of Oriented gradients dan Multiclass Support Vector Machine. *JATIKOM: Jurnal Aplikasi Dan Teori Ilmu Komputer*, 3(1), 1–7.
- Ye, T. (2024). *Head Detection Based on YOLOv8*. Github. <https://github.com/Owen718/Head-Detection-Yolov8>
- Yu, D., Su, K., Geng, X., & Wang, C. (2019). *A Context-and-Spatial Aware Network for Multi-Person Pose Estimation*. <http://arxiv.org/abs/1905.05355>
- Zhou, S., Chen, B., Fu, E. S., & Yan, H. (2023). Computer vision meets microfluidics: a label-free method for high-throughput cell analysis. *Microsystems & Nanoengineering*, 9(1), 116. <https://doi.org/10.1038/s41378-023-00562-8>
- Zhu, X., & Goldberg, A. (2009). *Introduction to semi-supervised learning*. Morgan & Claypool Publishers.
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2021). A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE*, 109(1), 43–76. <https://doi.org/10.1109/JPROC.2020.3004555>
- Zou, Z., Chen, K., Shi, Z., Guo, Y., & Ye, J. (2023). Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, 111(3), 257–276. <https://doi.org/10.1109/JPROC.2023.3238524>