ENS Paris-Saclay, MVA 2022-2023

Using Error Level Analysis to remove Underspecification

Jérémie Dentan, École Polytechnique, France

March 17th, 2023

This presentation comes with a technical report and a GitHub repository. It presents the work of the author on a data competition.

- *The competition: https://challengedata.ens.fr/participants/challenges/95/*
- *The technical report: http://dx.doi.org/10.13140/RG.2.2.25127.21925*
- *The GitHub repository: https://github.com/DentanJeremie/age-underspecification.git*

Content

Content

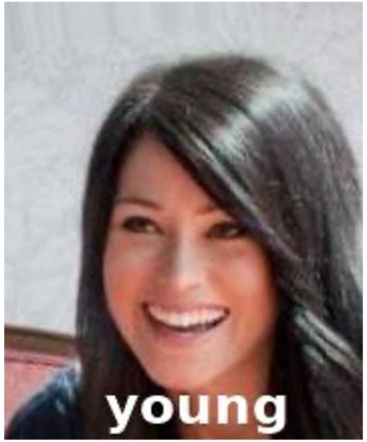
I- Introduction: The challenge

Huge distribution shift between the train and test set

The task: predict age from ambiguous data

Train

Train



AYTY

AYTO

AOTO

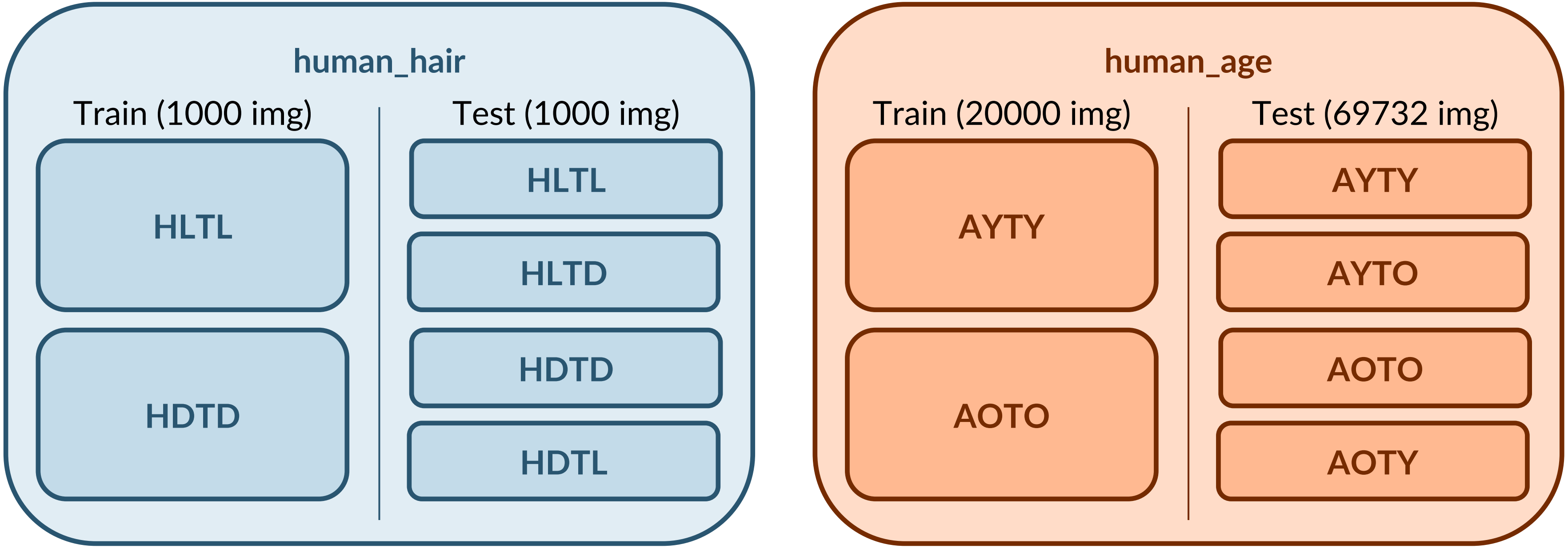
AOTY

Test

A=Age, T=text, Y=Young, O=Old

I- Introduction: The datasets

Two datasets: 2000 + 89732 JPEG-RGB images, size 178x218



HLTL, HDTD, HLTD, HDTL are que equivalent of AYTY, AOTO, AYTO, AOTY, replacing the Age by the Hair, and Old/Young by Light/Dark

I- Introduction: The rules

Some rules were there to make the challenge harder.

No manual annotation	→	• We need machine learning techniques
No other data	→	• We cannot scrap data to train an age classifier
No pretrained model, except the ones trained on ImageNet	→	• We cannot used pretrained age classifier • Difficult to use out-of-the box OCR, often pretrained

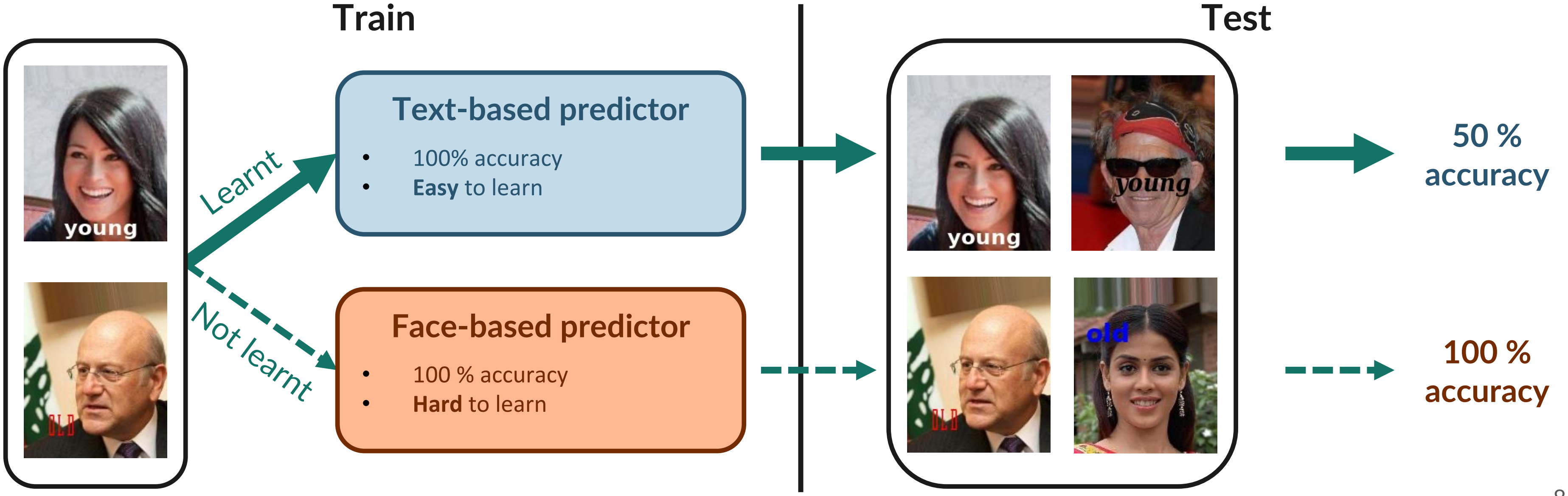
Content

Ref: [2, 4, 5, 8, 19]

Synonym: ambiguity of the data

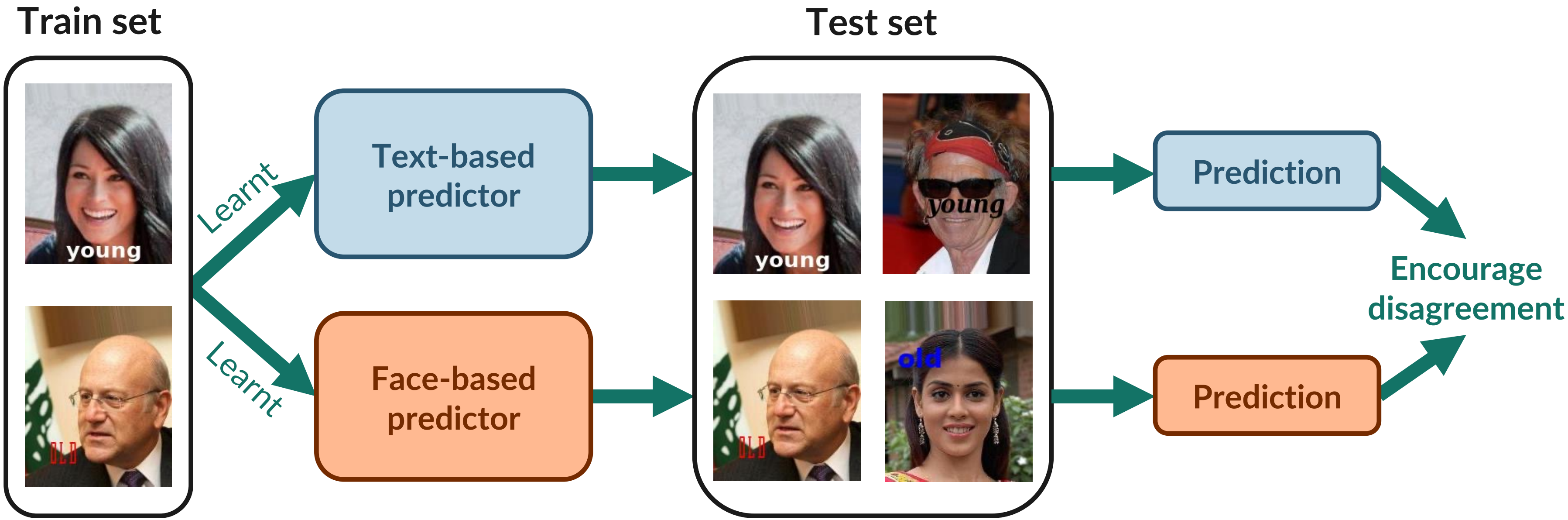
II- Underspecification: What is it?

During the training, the simplest predictor is learnt, however it is not always the one that have the best performance on the test set



II- Underspecification: One solution: DivDis

DivDis is an architecture that uses both the train and test set during the training, encouraging disagreement between the two predictors it learns.



II- Underspecification: Discussion on DivDis

DivDis is hard to apply in real life, because we need an expert to understand where the ambiguity comes from and how to characterize it.

Challenges with the architecture:

- Need to be sure there is a distribution shift between two datasets
- Need to choose which predictor to deploy on real data

→ Requires an expert to **understand** the ambiguity and **characterize it**

Approach 1

- Use the DivDis architecture
- Characterize the ambiguity *a posteriori*
- Choose one of the two predictor

Approach 2

- Characterize the ambiguity *a priori*
- Remove it
- Train a single predictor

Content

Ref: [12] (DivDis) ; [6] (ImageNet) ; [9] (ResNet)

III- First approach: DivDis with a pretrained head

To adapt the DivDis architecture to our problem, we wanted to pretrain one of the two heads to read the text on the images.

human_hair + human_age

4-label classification

- ResNet50 architecture
- Frozen backbone
- 20% images for validation
- Data replication to have balanced dataset, with: (1) Gaussian noise (2) Rolling color channels (3) Reverting lightness

Merging the two datasets, the text was supposed to be no longer correlated with the image

This did not work well on the unlabeled image, so we abandoned this approach

Content

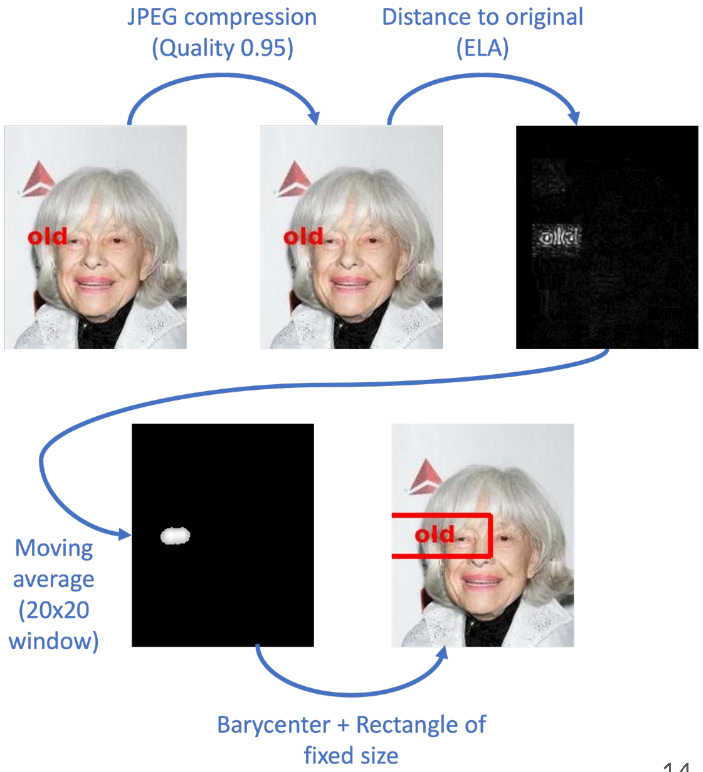
III- Second approach: Removing the text

We adapted a forensic technique called Error Level Analysis (ELA) to remove the text from the image, with perfectly worked in 91% of the cases.

Some examples of detections:

- Image 1 and 2: correct detection, as in 91% of cases. The detection was difficult: presence of another text in the image or low contrast.
- Image 3 and 4: partially incorrect detection or fully incorrect detection, as respectively in 4% and 5% of cases.

The accuracy rates are manually evaluated using 200 random images.

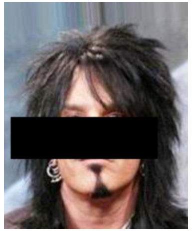


Ref: [10, 20] (Age classification pipelines) ; [9, 22] (ResNet50, VGG19) ; [11] (Adam)

III- Second approach: Training the classifier

After having removed the text, we trained a binary classifier, leading to a final accuracy of 73% (vs. 64% for the DivDis baseline).

data



Binary classification

- ResNet50 and VGG19 architectures
- Adam Optimizer
- Binary cross-entropy
- 20% images for validation
- Gradual unfreezing across 13 epochs
- Early stopping and memorization of the best validation accuracy
- Data replication with gaussian noise

Some metrics:

- 73% accuracy on test set (vs. 64% for the DivDis baseline).
- CPU for text removal: Intel Xeon W-1290P 3.70GHz, 20 virtual cores: 455sec, i.e. about 10sec/img/proc
- GPU for age prediction: NVIDIA GeForce RTX 3090 24Go: 760sec/epoch, i.e. about 2h45 total training time per model

Content

IV- Conclusion

- We have tried **two different approaches** to fight the ambiguity of the data at hand
- Removing the ambiguity ***a priori*** does not require more domain knowledge than doing it ***a posteriori*** as with DivDis
- We showed that simple **forensic techniques** can prove to be very efficient to detect some patterns to remove the ambiguity
- We obtained a satisfying **final accuracy of 73%** (vs. 64% for the DivDis baseline)

References

[1] 2023. Tesseract OCR. https://github.com/tesseract-ocr/tesseract original-date: 2014-08-12T18:04:59Z.

[2] Dario Amodei, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané. 2016. Concrete Problems in AI Safety. https://doi.org/10.48550/arXiv.1606.06565 arXiv:1606.06565 [cs].

[3] Martin Arjovsky, Léon Bottou, Ishaan Gulrajani, and David Lopez-Paz. 2020. Invariant Risk Minimization. https://doi.org/10.48550/arXiv.1907.02893 arXiv:1907.02893 [cs, stat].

[4] Devansh Arpit, Stanisław Jastrzębski, Nicolas Ballas, David Krueger, Emmanuel Bengio, Maxinder S. Kanwal, Tegan Maharaj, Asja Fischer, Aaron Courville, Yoshua Bengio, and Simon Lacoste-Julien. 2017. A Closer Look at Memorization in Deep Networks. https://doi.org/10.48550/arXiv.1706.05394 arXiv:1706.05394 [cs, stat].

[5] Alexander D'Amour, Katherine Heller, Dan Moldovan, Ben Adlam, Babak Alipanahi, Alex Beutel, Christina Chen, Jonathan Deaton, Jacob Eisenstein, Matthew D. Hoffman, Farhad Hormozdiari, Neil Houlsby, Shaobo Hou, Ghassen Jerfel, Alan Karthikesalingam, Mario Lucic, Yian Ma, Cory McLean, Diana Mincu, Akinori Mitani, Andrea Montanari, Zachary Nado, Vivek Natarajan, Christopher Nielson, Thomas F. Osborne, Rajiv Raman, Kim Ramasamy, Rory Sayres, Jessica Schrouff, Martin Seneviratne, Shannon Sequeira, Harini Suresh, Victor Veitch, Max Vladymyrov, Xuezhi Wang, Kellie Webster, Steve Yadlowsky, Taedong Yun, Xiaohua Zhai, and D. Sculley. 2020. Underspecification Presents Challenges for Credibility in Modern Machine Learning. https://doi.org/10.48550/arXiv.2011.03395 arXiv:2011.03395 [cs, stat].

[6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255. https://doi.org/10.1109/CVPR.2009.5206848 ISSN: 1063-6919.

[7] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. 2016. Domain-Adversarial Training of Neural Networks. https://doi.org/10.48550/arXiv.1505.07818 arXiv:1505.07818 [cs, stat].

[8] Suriya Gunasekar, Jason Lee, Daniel Soudry, and Nathan Srebro. 2019. Implicit Bias of Gradient Descent on Linear Convolutional Networks. https://doi.org/10.48550/arXiv.1806.00468 arXiv:1806.00468 [cs, stat].

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. Deep Residual Learning for Image Recognition. http://arxiv.org/abs/1512.03385 arXiv:1512.03385 [cs].

[10] Shakediel Hiba and Yosi Keller. 2021. Hierarchical Attention-based Age Estimation and Bias Estimation. http://arxiv.org/abs/2103.09882 arXiv:2103.09882 [cs] version: 1.

[11] Diederik P. Kingma and Jimmy Ba. 2017. Adam: A Method for Stochastic Optimization. https://doi.org/10.48550/arXiv.1412.6980 arXiv:1412.6980 [cs].

[12] Yoonho Lee, Huaxiu Yao, and Chelsea Finn. 2022. Diversify and Disambiguate: Learning From Underspecified Data. https://doi.org/10.48550/arXiv.2202.03418 arXiv:2202.03418 [cs, stat].

References

[13] Minghao Li, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, and Furu Wei. 2022. TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models. https://doi.org/10.48550/arXiv.2109.10282 arXiv:2109.10282 [cs].

[14] Junyang Lin, Xuancheng Ren, Yichang Zhang, Gao Liu, Peng Wang, An Yang, and Chang Zhou. 2022. Transferring General Multimodal Pretrained Models to Text Recognition. https://doi.org/10.48550/arXiv.2212.09297 arXiv:2212.09297 [cs].

[15] Evan Zheran Liu, Behzad Haghgoo, Annie S. Chen, Aditi Raghunathan, Pang Wei Koh, Shiori Sagawa, Percy Liang, and Chelsea Finn. 2021. Just Train Twice: Improving Group Robustness without Training Group Information. https://doi.org/10.48550/arXiv.2107.09044 arXiv:2107.09044 [cs, stat].

[16] Jörg Meyer. 2012. *Forensische Datenanalyse : dolose Handlungen im Unternehmen erkennen und aufdecken*. Erich Schmidt. Google-Books-ID: V96iMAEACAAJ.

[17] Junhyun Nam, Hyuntak Cha, Sungsoo Ahn, Jaeho Lee, and Jinwoo Shin. 2020. Learning from Failure: Training Debiased Classifier from Biased Classifier. https://doi.org/10.48550/arXiv.2007.02561 arXiv:2007.02561 [cs, stat].

[18] Cao Hong Nga, Khai-Thinh Nguyen, Nghi C. Tran, and Jia-Ching Wang. 2020. Transfer Learning for Gender and Age Prediction. In *2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*. 1–2. https://doi.org/10.1109/ICCE-Taiwan49838.2020.9258347 ISSN: 2575-8284.

[19] Luke Oakden-Rayner, Jared Dunnmon, Gustavo Carneiro, and Christopher Ré. 2019. Hidden Stratification Causes Clinically Meaningful Failures in Machine Learning for Medical Imaging. https://doi.org/10.48550/arXiv.1909.12475 arXiv:1909.12475 [cs, stat].

[20] Zakariya Qawaqneh, Arafat Abu Mallouh, and Buket D. Barkana. 2017. Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model. http://arxiv.org/abs/1709.01664 arXiv:1709.01664 [cs].

[21] Shiori Sagawa, Pang Wei Koh, Tatsunori B. Hashimoto, and Percy Liang. 2020. Distributionally Robust Neural Networks for Group Shifts: On the Importance of Regularization for Worst-Case Generalization. https://doi.org/10.48550/arXiv.1911.08731 arXiv:1911.08731 [cs, stat].

[22] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. http://arxiv.org/abs/1409.1556 arXiv:1409.1556 [cs].

[23] A. Skodras, C. Christopoulos, and T. Ebrahimi. 2001. The JPEG 2000 still image compression standard. *IEEE Signal Processing Magazine* 18, 5 (Sept. 2001), 36–58. https://doi.org/10.1109/79.952804 Conference Name: IEEE Signal Processing Magazine.

[24] Ida Sudiatmika, Fathur Rahman, Trisno Trisno, and Suyoto Suyoto. 2018. Image forgery detection using error level analysis and deep learning. *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 17 (Aug. 2018), 653. https://doi.org/10.12928/telkomnika.v17i2.8976

[25] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. 2014. Deep Domain Confusion: Maximizing for Domain Invariance. https://doi.org/10.48550/arXiv.1412.3474 arXiv:1412.3474 [cs].