



# Catastrophic Interference in Disguised Face Recognition

Parichehr B. Ardakani<sup>1</sup>(✉), Diego Velazquez<sup>1</sup>, Josep M. Gonfaus<sup>2</sup>,  
Pau Rodríguez<sup>3</sup>, F. Xavier Roca<sup>1</sup>, and Jordi Gonzàlez<sup>1</sup>

<sup>1</sup> Computer Vision Center, Univ. Autònoma de Barcelona, 08193 Bellaterra, Spain  
[pbehjati@cvc.uab.es](mailto:pbehjati@cvc.uab.es)

<sup>2</sup> Visual Tagging Services, Parc de Recerca UAB, 08193 Bellaterra, Spain

<sup>3</sup> Element AI, Montreal 6650, Canada

**Abstract.** It is commonly known the natural tendency of artificial neural networks to completely and abruptly forget previously known information when learning new information. We explore this behaviour in the context of Face Verification on the recently proposed Disguised Faces in the Wild dataset (DFW). We empirically evaluate several commonly used DCNN architectures on Face Recognition and distill some insights about the effect of sequential learning on distinct identities from different datasets, showing that the catastrophic forgetness phenomenon is present even in feature embeddings fine-tuned on different tasks from the original domain.

**Keywords:** Neural network forgetness · Face recognition · Disguised Faces

## 1 Introduction

Deep Convolutional Neural Networks (DCNNs) have achieved remarkable success in various cognitive applications such as image recognition, facial detection, signal processing, on supervised, unsupervised and reinforcement learning tasks through feature representations at successively higher, more abstract layers. Computational complexity and the time needed to train large networks is one of the major challenges for convolutional networks. It is common to pretrain a DCNN on a large dataset and then use the trained network as an initialization or as a fixed feature extractor for a particular application [24]. A major downside of such DCNNs is the inability to retain previous knowledge while learning new information. This problem is called Catastrophic forgetting.

Catastrophic forgetting is a term, often used in connectionist literature, to describe a common problem with many traditional artificial neural network models. It refers to forgetting what has been learned upon learning new or different

---

P. B. Ardakani and D. Velazquez—Both authors contributed equally to this work.

information. For instance, when a network is first trained to convergence on one task, and then trained on a second task, it forgets how to perform the former.

There are some approaches to improve the performance of models when learning new information that benefit from previously learned information, for example fine-tuning [6], where the old task parameters are adjusted to adapt to a new task. Other approaches well known are feature extraction [5] where the parameters of the old network are unchanged and the parameters of the outputs of one or more layers are used to extract features for the new task. There is also a paradigm called joint train [4] in which parameters of old and new tasks are trained together to minimize the loss in all tasks.

Overcoming the problem of catastrophic forgetting is an important step. Some methods have already been developed to overcome this problem [8, 17, 27]. But even with these and other methods, the problem of catastrophic forgetting is still a key problem within the Artificial Intelligence (AI) community and it is time to move towards algorithms that can learn multiple tasks over time [25].

The novel approaches have been developed specifically for the tasks of visual face recognition and verification in order to boost performance on public datasets such as Labeled Faces in the wild (LFW) [11]. However, the performance on completely unconstrained datasets like Youtube Face (YTF) [30], and UMDFaces [1] remains subpar at low false alarm rates. These datasets contain significant variations in illumination, pose, expression, aging and tend to have low resolution and clutter filled images. This indicates that the problem of face recognition is far from *solved*. The recently announced Disguised Faces in the Wild (DFW) dataset aims to study another covariate of the face verification pipeline - *disguises*.

Disguise and impersonation are part of a sub-field of face recognition where the subjects are non-cooperative and are actively trying to deceive the system. A disguise involves both intentional or unintentional changes on a face through which one can either obfuscate his/her identity. This means that the subject is trying to adopt a new identity in order to hide his/her own. A subject might impersonate someone else's identity. Obfuscation increases the inter-class variations whereas impersonation reduces the inter-class dissimilarity, thereby affecting face recognition/verification task and making it non-trivial. This is a very challenging face verification problem and has not been studied in a comprehensive way, primarily due to the unavailability of such a dataset. The aim of a face verification system in such cases is to identify a given subject under varying disguises while rejecting impostors trying to look like the subject of interest in an uncontrolled setting. From the point of view of an automated computer vision method, it is important to extract rich face features in-order to distinguish among the identities and verify them correctly.

In this paper, we explore catastrophic behaviour in the context of Face Verification on the DFW dataset. We empirically evaluate several commonly used DCNN architectures on Face Recognition and distill some insights about the effect of sequential learning on distinct identities from different datasets which are explained in the following sections.

## 2 Related Work

In this section we briefly review some recent related work and proposed methods on face recognition/verification and catastrophic forgetting.

Disguised faces recognition focuses on recognizing the identity of disguised faces and impersonators. There is limited research focus on this topic. MiRA-Face [31] uses two CNNs networks one for aligned input and the other for unaligned input to perform generic face recognition. Then, Principal Component Analysis (PCA) is used to find the transformation matrix for face recognition adaptation. Another work is Deep Disguise Recognizer (DDRNET) [14] uses an Inception Network along with Center loss [29] followed by classification using a similarity metric. DisguisedNet [28] proposed a Siamese-based approach using the pretrained VGG-Face [20] and after that, cosine similarity is applied for performing classification of the learned features. AEFRL [26], performs face detection and alignment on the input images using Multi-task Cascaded Convolutional Networks(MT-CNN) [32] followed by horizontal flipping. An ensemble of five networks is used to obtain features for original and flipped images. The concatenation of these features are used to perform classification using cosine similarity. UMDNets [2] is another work which uses All-in-One [21] to align the images using facial landmarks. They performs feature extraction using two networks, followed by independent score computation. Then, classification is performed by averaging the scores obtained via the two feature sets. Table 1 provides a list of the proposed approaches on DFW dataset for face verification.

The problem of catastrophic forgetting is a big issue in machine learning and artificial intelligence if our goal is to build a system that learns through time, and is able to deal with more than a single problem. According to [18], without this capability we will not be able to build truly intelligent systems, we can only create models that solve isolated problems in a specific domain. There are some recent works that tried to overcome this problem, e.g., domain adaptation that uses the knowledge learned to solve one task and transfers it to help learning another, but those two tasks have to be related. This approach was used in [12] to avoid the problem of catastrophic forgetting, in order to do so they use two properties. The first property was to keep the decision boundary unchanged and the second one, was that the feature extractor from the source data by the target network should be present in a position close to the features extracted from the source data by the source network. As was shown in the experiments, by keeping the decision boundaries unchanged new classes cannot be learned, making this approach unable to deal with related tasks that present a different number of classes. Early attempts to alleviate catastrophic forgetting often consists of a memory system that store previous data and replays the sampled old examples with the new data [22], and similar approaches are still used today [16]. [23] learns a generative model to capture the data distribution of previous tasks, and both generated samples and real samples from the current task are used to train the new model so that the forgetting can be alleviated for continual learning.

In our work, we will show that the intrinsic forgetness property of neural networks not only present when performing classification on new problems but also when extracting features even for tasks whose domain is the same.

**Table 1.** Different approaches to face verification

Model	Brief description
AEFRL	MTCNN + 4 Networks for feature extraction + Cosine distance
DDRNET	Weighted linear combination of ensemble of 3 CNNs
DisguiseNet	VGG siamese architecture + Weighted Loss + Cosine Distance
MiRA-Face	MTCNN + RSA + Ensemble of CNNs Text follows
UMDNets	All-in-One + average across scores obtained by 2 networks



**Fig. 1.** Some example images of Disguised Faces in the wild (DFW) dataset. The dataset contains four kinds of images: normal, validation, disguise and impersonator (figure taken from [15])

### 3 Methodology

In this section we describe the DFW dataset [15] along with the evaluation protocols it presents. We comment on the different architectures we use for our experiments and explain the identity overlap between the datasets used for training and the DFW test set.

#### 3.1 Dataset

In our experiments we evaluate the performance of different models in the face verification task using the DFW dataset. This dataset has been created to address the issue of disguised and imposter faces in the wild. The dataset consists of 11,157 images belonging to 1000 different subjects. Images pertaining

to 400 subjects form the training set, while the remaining 600 subjects constitute the test set. Each subject has at least five face images, and can have four types of images: (i) *normal*, (ii) *validation*, (iii) *disguised* and (iv) *impersonator*.

Normal images correspond to non-disguised frontal face images. Validation images are used to generate a non-disguised pair within a subject. Disguised images correspond to a face image of the same subject having intentional or unintentional disguise. Impersonator images correspond to a face image of individuals who intentionally or unintentionally look similar to a different subject. An example of each is shown in Fig. 1.

Three verification protocols have been provided with the DFW dataset to understand and evaluate the effect of disguises on face recognition.

- **Protocol-1 (Impersonation)** evaluates the capacity of the system to differentiate genuine users from impersonators. Genuine pairs for this protocol are created by combining a genuine image and a validation image from the same subject. Impostors pairs are created by combining impersonator images with normal, validation and disguised images from the same subject. This protocol is made up of 25,046 possible pairs.
- **Protocol-2 (Obfuscation)** evaluates the robustness of the system when it comes to detecting when a subject is unintentionally or intentionally trying to hide his identity. The genuine set for this protocol comprises pairs formed by (normal, validation), (validation, disguise) and (disguise<sub>1</sub>, disguise<sub>2</sub>) images from the same subject. Where disguised<sub>n</sub> corresponds to the  $n^{th}$  disguised image of a subject. Impostor pairs are generated by creating cross-subject pairs, combining normal, validation and disguised images of one subject with their counterpart from another subject. This protocol consists of 9,041,283 possible pairs.
- **Protocol-3 (Overall Performance)** is a the combination of the previous two and evaluates the overall performance of the system. A valid genuine or impostor pair for this protocol can be any genuine or impostor pair from protocols 1 and 2. This protocol comprises 9,066,329 possible pairs.

### 3.2 Neural Network Architectures

In order to carry out the experiments we used three neural network architectures: (i) VGG-Face [20] (ii) ResNet-50 [9] (iii) Se-ResNet-50 [10]. The training and testing details will be explained in the following section.

**VGG-Face.** In our first experiment, we use a pretrained implementation of the VGG-Face CNN which is one of the top performing deep learning models for face recognition, this will act as our baseline for the rest of the experiments. The network was trained on the VGG-FACE dataset [20].

**Table 2.** Datasets used for the training of each model. The last column refers to the number of different identities present in the training set of each dataset that can also be found in the DFW test set.

Dataset	VGG	Resnet50	Resnet50-ft	Senet	Overlapping identities
VGG-Face	✓				203 (33%)
VGG-Face2		✓	✓	✓	122 (20%)
MS-Celeb-1M			✓	✓	348 (58%)
DFW (non-overlapping)					143 (24%)

**ResNet-50.** In the next experiment, we use two residual networks for our face verification system, concretely two Resnet-50. One network is trained on MS-Celeb-1M [7] and then fine-tuned VGG-Face2 [3], while the other one is just trained on VGG-Face2. The architecture comprises 50 convolutional layers followed by a fully connected layer of dimension 2048.

**Se-ResNet-50.** Lastly we use a pretrained Se-Resnet-50 in our last experiment. This network is trained on MS-Celeb-1M. The only difference between the architecture of this model and ResNet-50 is that ‘Squeeze-and-Excitation’ (SE) block is added to the convolutional layers of the ResNet-50 followed by an embedding of 256 dimension. SE block can be used with any standard architecture. The SE block tries to use global information to selectively emphasize informative features and suppress less useful once.

### 3.3 Dataset Overlap

The datasets that were used to pretrain the models we are evaluating present overlapping identities with the DFW test set. Despite containing the same identities, the face images do not need to be the same. Studying how each architecture performs when evaluated on these identities will provide us with insight into the ability of statistical models to retain and generalize previously acquired knowledge when fitting a new distribution. Table 2 shows which dataset was used to train each of the models we evaluate. Note that there are also identities from DFW that overlap in more than one dataset:  $VGG-Face \cap MS-Celeb-1M = 145$ , and  $VGG-Face2 \cap MS-Celeb-1M = 71$ .

## 4 Experiments and Results

In this section we present the different experiments and results obtained on every DFW protocol over every overlapping set. We also present some hard examples and an embedding visualization.

**Table 3.** Verification accuracy (%) of the different approaches and our results (last 4 rows). Models are evaluated on protocol-1 (P1), protocol-2 (P2) and protocol-3 (P3). *Senet + Resnet50-ft* represents an embedding of these two models

Algorithm	GAR-P1		GAR-P2		GAR-P3	
	1% FAR	0.1% FAR	1% FAR	0.1% FAR	1% FAR	0.1% FAR
Baseline (VGG-Face)	55.29	28.91	34.32	17.58	36.25	19.35
AEFRL	96.08	57.64	87.82	77.06	87.90	75.54
DDRNET	84.20	51.26	71.04	49.28	71.43	49.08
MIRA-Face	95.46	51.09	90.65	80.56	90.62	79.26
UMDNets	94.28	53.27	86.62	74.69	86.75	72.90
DisguiseNet	1.34 <sup>a</sup>	1.34 <sup>b</sup>	66.32	28.99	60.89	23.25
Resnet50	81.18	49.92	75.63	55.16	75.92	54.26
Resnet50-ft	83.70	53.45	77.91	58.37	78.00	56.98
Senet	86.72	50.92	78.93	60.39	79.07	58.92
Senet+Resnet50-ft	86.89	55.63	80.71	63.02	80.89	61.12

<sup>a</sup>GAR@0.95%FAR

<sup>b</sup>The smallest FAR value is 0.95% for DisguiseNet

#### 4.1 Performance on DFW

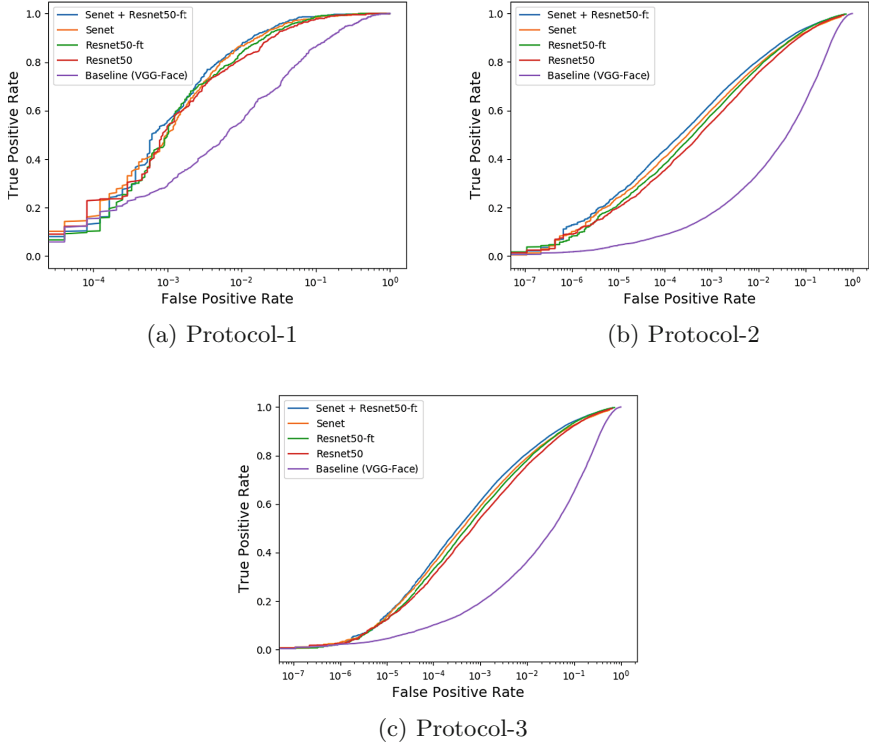
First, we evaluate and compare the different models on the standard dataset. We use the Genuine Acceptance Rate (GAR) at False Acceptance Rate of 1% and 0.1% (FAR), as defined in the original paper [15]. Table 3 shows the results obtained by several algorithms in each of the DFW evaluation protocols. The top performing methods do so well because they use models pretrained with over 5M images and fine-tune them on the DFW dataset for the face verification task.

Figure 2 shows the results of our experiments on each DFW protocol. It is clear that the models obtain competitive results despite none of them being specifically trained for this task, or fine-tuned in the DFW training set. This is, of course, due to the aforementioned identity overlap and the high capacity of the models used.

#### 4.2 Dataset Overlapping Study

As presented on Sect. 3.3 the datasets that were used to pretrain the models have overlapping identities with the DFW test set.

Model performance can vary significantly when evaluated on different subsets of the data, mainly due to the difficulty of the image pairs from each subset. Despite this, the overall performance is directly correlated with the model capacity and the quantity of images seen during training. Table 4 presents the performance of every evaluated model across different overlapping sets of identities. Scores on overlapping sets of identities *seen* by the architecture during training are presented in bold. It is easy to understand that the models will perform better on these subsets of the data.



**Fig. 2.** ROC Curves for every evaluated model on each DFW protocol

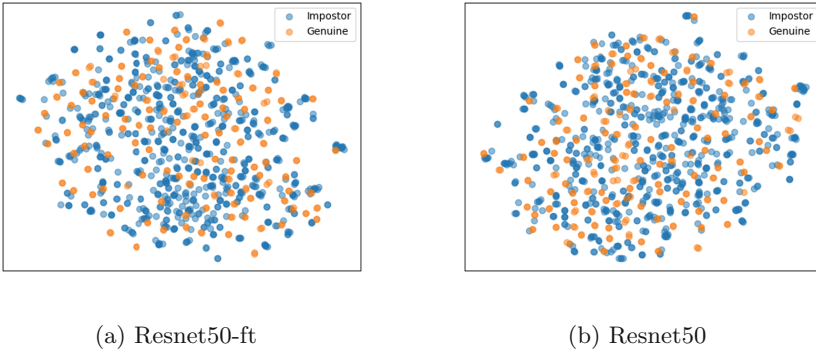
Catastrophic forgetting in neural networks occurs because of the stability-plasticity dilemma [13]. The model requires sufficient plasticity to acquire new tasks, but large weight changes will cause forgetting by distributing previously learned representations. A concrete example of catastrophic forgetting is when a network is training on new tasks or categories, a neural network tends to forget the information learned in the previous trained tasks from different domains. This usually means a new task will likely override the weights that have been learned in the past, and thus degrade the model performance for the past tasks. In this work we show that as the domain of two task remains unchanged, the weight changes are small, therefore the improvement ratio of the fine-tuned ResNet over the original model (*Resnet50-ft* vs *Resnet50*) remains constant ( $\sim 3\%$ ) across different overlapping sets. This effect indicates that the fine-tuned network is not able to retain specific knowledge from the first distribution it was trained on (the Ms-Celeb-1M dataset). If this were not the case, the fine-tuned network would perform much better than the original model on this overlapping set. Therefore, the overall improvement seems to arise solely from the increase in *seen* images.



Due to the intrinsic forgetness property of statistical models learning multiple task from mutually exclusive domains, without forgetting all but one of them, is unfeasible. However, this experiment shows that even when the domain of the learned tasks are the same, the catastrophic forgetness problem persists. Therefore, the forgetness problem seems to not only affect the fully connected layers acting as classifiers, but also the deepest layers in charge of feature extraction.

**Table 4.** Performance (GAR@1%FAR) of every evaluated model across different overlapping sets of subjects. The scores in bold indicate the performance of the model on identities *seen* during training

Overlapping set	VGG	Resnet50	Resnet50-ft	Senet
VGG-Face	<b>39.63</b>	73.79	74.71	76.25
VGG-Face2	36.59	<b>77.70</b>	<b>80.65</b>	<b>84.70</b>
Ms-Celeb-1M	35.34	73.97	<b>75.64</b>	<b>76.66</b>
VGG-Face $\cap$ Ms-Celeb-1M	<b>41.28</b>	76.85	<b>80.03</b>	<b>79.94</b>
VGG-Face2 $\cap$ Ms-Celeb-1M	28.59	<b>77.64</b>	<b>79.70</b>	<b>81.16</b>
None Overlapping	35.14	75.84	77.60	78.79
All sets	36.25	75.92	78.00	79.07



**Fig. 3.** Embedding representation of Genuine and impostor subjects that overlap with the MS-Celeb-1M dataset

### 4.3 Face Embedding Representation

The forgetness property can also be analyzed by projecting the face image embeddings of impostors and genuine subjects into a 2d space using t-SNE [19]. All the face images were created by padding the provided face coordinates and

resizing the resulting bounding box maintaining aspect ratio with shorter side of 256 and then center cropped to  $224 \times 224$ .

Note on Fig. 3, how the embeddings of both Resnet50 architectures struggle similarly to separate impostors from genuine subjects further evidencing our hypothesis stating that the fine-tuned architecture has forgotten the faces it was originally trained on. The apparently random distribution of both embeddings also demonstrate the high level of complexity that the face verification task represents.



**Fig. 4.** The pairs consistently misclassified by every model. These are genuine pairs that were labeled as an impostor pair (false negatives). All the pairs have been extracted from Protocol-3, since it is the one that comprises every possible pair.

#### 4.4 Visualizing Hard Examples

To shed some light into the difficulty of the face verification task in the DFW dataset, we show some examples of pairs commonly misclassified by every architecture on Protocol-3. Figure 4 shows some hard genuine pairs. Note how often the misclassified genuine pairs represent drastic changes in face structure, pose and texture. This makes it notoriously hard, even for humans, to correctly classify these pairs.

## 5 Conclusions

In this study we show that the intrinsic forgetness property of neural networks is not only present when performing classification but also when extracting features for similar tasks sharing the same domain. After the fine-tuning process, even powerful architectures like Resnet50 will fail to remember the distribution they first learned.

In our experiments, we observe that the model that has been pretrained on MS-Celeb-1M and then fine-tuned on VGG-dataset-2, has a relatively constant improvement across different overlapping subsets of identities. This behaviour indicates that the model has forgotten some specifics about the previously fitted distribution to accommodate a new one. The consistent gain in accuracy across different overlapping subsets is solely due to the larger amount of *seen* images.

**Acknowledgements.** Authors acknowledge the support of the Spanish project TIN2015-65464-R (MINECO FEDER), the 2016FLB 01163 grant of Generalitat de Catalunya, and the COST Action IC1307 iV&L Net. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of a Tesla K40 GPU and a GTX TITAN GPU, used for this research.

## References

1. Bansal, A., Nanduri, A., Castillo, C.D., Ranjan, R., Chellappa, R.: UMDFaces: an annotated face dataset for training deep networks. In: 2017 IEEE International Joint Conference on Biometrics (IJCB), pp. 464–473. IEEE (2017)
2. Bansal, A., Ranjan, R., Castillo, C.D., Chellappa, R.: Deep features for recognizing disguised faces in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 10–16 (2018)
3. Cao, Q., Shen, L., Xie, W., Parkhi, O.M., Zisserman, A.: VGGFace2: a dataset for recognising faces across pose and age. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), pp. 67–74. IEEE (2018)
4. Caruana, R.: Multitask learning. *Mach. Learn.* **28**(1), 41–75 (1997)
5. Donahue, J., et al.: DeCAF: a deep convolutional activation feature for generic visual recognition. In: International Conference on Machine Learning, pp. 647–655 (2014)
6. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv preprint [arXiv:1311.2524](https://arxiv.org/abs/1311.2524) (2014)
7. Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J.: MS-Celeb-1M: a dataset and benchmark for large-scale face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9907, pp. 87–102. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46487-9\\_6](https://doi.org/10.1007/978-3-319-46487-9_6)
8. Gutstein, S., Stump, E.: Reduction of catastrophic forgetting with transfer learning and ternary output codes. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–8. IEEE (2015)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
10. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)
11. Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. In: Workshop on Faces in ‘Real-Life’ Images: Detection, Alignment, and Recognition (2008)
12. Jung, H., Ju, J., Jung, M., Kim, J.: Less-forgetting learning in deep neural networks. arXiv preprint [arXiv:1607.00122](https://arxiv.org/abs/1607.00122) (2016)
13. Kemker, R., McClure, M., Abitino, A., Hayes, T.L., Kanan, C.: Measuring catastrophic forgetting in neural networks. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
14. Kohli, N., Yadav, D., Noore, A.: Face verification with disguise variations via deep disguise recognizer. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 17–24 (2018)

15. Kushwaha, V., Singh, M., Singh, R., Vatsa, M., Ratha, N., Chellappa, R.: Disguised faces in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1–9 (2018)
16. Lesort, T., Caselles-Dupré, H., Garcia-Ortiz, M., Stoian, A., Filliat, D.: Generative models from the perspective of continual learning. arXiv preprint [arXiv:1812.09111](https://arxiv.org/abs/1812.09111) (2018)
17. Li, Z., Hoiem, D.: Learning without forgetting. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(12), 2935–2947 (2018)
18. Liu, B.: Lifelong machine learning: a paradigm for continuous learning. *Front. Comput. Sci.* **11**(3), 359–361 (2017)
19. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**(Nov), 2579–2605 (2008)
20. Parkhi, O.M., Vedaldi, A., Zisserman, A.: Deep face recognition. In: British Machine Vision Conference (2015)
21. Ranjan, R., Sankaranarayanan, S., Castillo, C.D., Chellappa, R.: An all-in-one convolutional neural network for face analysis. In: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pp. 17–24. IEEE (2017)
22. Robins, A.: Catastrophic forgetting, rehearsal and pseudorehearsal. *Connection Sci.* **7**(2), 123–146 (1995)
23. Rusu, A.A., et al.: Progressive neural networks. arXiv preprint [arXiv:1606.04671](https://arxiv.org/abs/1606.04671) (2016)
24. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S.: CNN features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 806–813 (2014)
25. Silver, D.L., Yang, Q., Li, L.: Lifelong machine learning systems: beyond learning algorithms. In: 2013 AAAI Spring Symposium Series (2013)
26. Smirnov, E., Melnikov, A., Oleinik, A., Ivanova, E., Kalinovskiy, I., Luckyanets, E.: Hard example mining with auxiliary embeddings. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 37–46 (2018)
27. Van Merriënboer, B., et al.: Blocks and fuel: frameworks for deep learning. arXiv preprint [arXiv:1506.00619](https://arxiv.org/abs/1506.00619) (2015)
28. Vishwanath Peri, S., Dhall, A.: DisguiseNet: a contrastive approach for disguised face verification in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 25–31 (2018)
29. Wen, Y., Zhang, K., Li, Z., Qiao, Y.: A discriminative feature learning approach for deep face recognition. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 499–515. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46478-7\\_31](https://doi.org/10.1007/978-3-319-46478-7_31)
30. Wolf, L., Hassner, T., Maoz, I.: Face recognition in unconstrained videos with matched background similarity. *IEEE* (2011)
31. Zhang, K., Chang, Y.L., Hsu, W.: Deep disguised faces recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 32–36 (2018)
32. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **23**(10), 1499–1503 (2016)