

# Reidentyfikacja plam zabrudzeń

Jan Kwiatkowski, Alicja Turowska, Konrad Gieleta  
Kwiecień 2023

# Cel i analiza tematu

- Celem zadania jest stworzenie modelu służącego do reidentyfikacji plam zabrudzeń.
- Reidentyfikacja jest niezbędna w celu śledzenia poruszających się obiektów w sekwencji klatek.
- Reidentyfikację można sprowadzić do problemu znalezienia "najbliższego" innego zdjęcia.

# Cel i analiza tematu

- Reidentyfikacja plam zabrudzeń nie jest zbyt popularnym problemem, sama reidentyfikacja jest często poruszana w kontekście reidentyfikacji osób, gdzie do najczęściej wykorzystywanych metod należą konwolucyjne sieci neuronowe (CNN).
- Zaletą tych sieci jest automatyczna ekstrakcja cech obrazu, co odróżnia je od innych algorytmów rozpoznawania obrazu, które wymagają wiedzy eksperckiej z zakresu inżynierii cech.

# Cel i analiza tematu

Popularna funkcja kosztu: Triplet Loss

dostarczanie trójek wektorów osadzenia:

- zdjęcia będącego odniesieniem,
- zdjęcia "bliskiego",
- zdjęcia "dalekiego"

# Cel i analiza tematu

ZHENG, Zhedong; ZHENG, Liang; YANG, Yi. A discriminatively learned cnn embedding for person reidentification. *ACM transactions on multimedia computing, communications, and applications (TOMM)*, 2017, 14.1: 1-20.

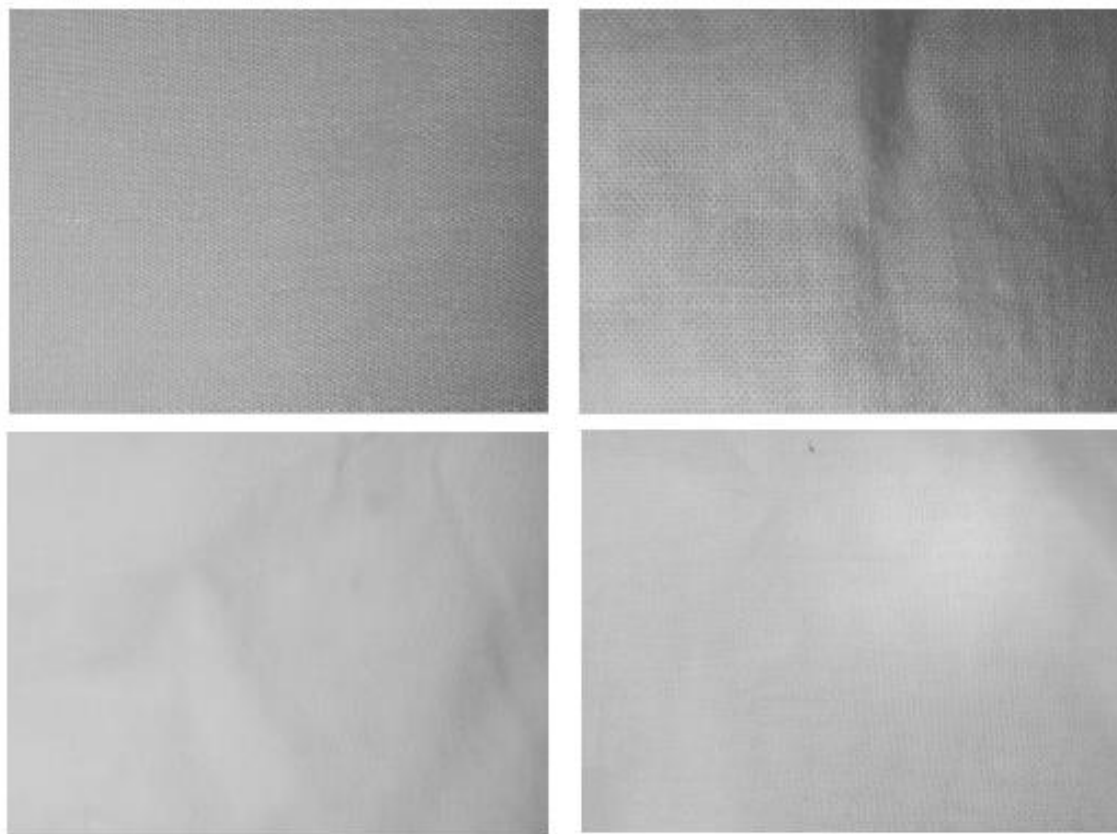
Inne podejście do problemu reidentyfikacji:

- Sieć syjamska (model identyfikacyjny i weryfikacyjny)
- dla pary obrazów szkoleniowych przewiduje tożsamość (model identyfikacyjny), oraz to czy należą do tej samej tożsamości, czyli kategorii (model weryfikacyjny)

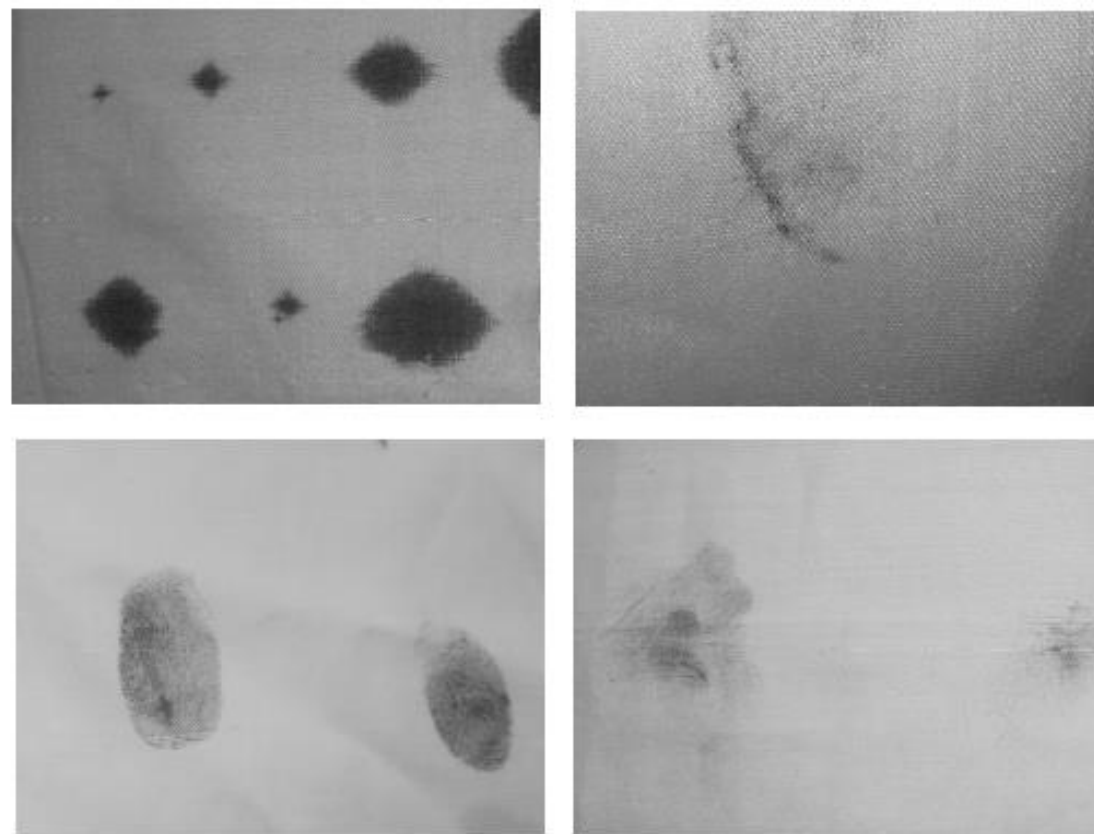
# Zbiory publiczne

- [TILDA](#):
  - nie dotyczy samych plam, co wszelakich defektów tkanin.
- [Fabric Stain Dataset](#):
  - dwie kategorie: tkaniny bez defektów oraz tkaniny z plamami
  - 466 zdjęć: 68 zdjęć tkanin bez zabrudzeń oraz 398 zdjęć tkanin z plamami atramentu, brudu, oraz oleju (bez podziału na podkategorie)
  - rozdzielczości: 1488x1984 (305 zdjęć) oraz 1984x1488 (161 zdjęć)
  - brak podziału na zbiór treningowy, walidacyjny i testowy

# Przykładowe zdjęcia ze zbioru Fabric Stain Dataset



Zdjęcia tkanin bez defektów



Zdjęcia tkanin z plamami

# Założenia projektu

- Zbiory:
  - publiczny zbiór Fabric Stain Dataset,
  - zbiór stworzony na potrzeby projektu,
- Modele:
  - ResNet50 z funkcją kosztu triplet loss,
  - VGG16
  - EfficientNet B7



# Resnet-50

- Z rodziny głębokich sieci konwolucyjnych - Residual Neural Network, rozwiązującej problem zanikania gradientów,
- 16 bloków rezydualnych zawierających po 3 warstwy konwolucyjne, po których jest funkcja aktywacji, do danych wyjściowych bloku dodawane są oryginalne dane wejściowe,
- Charakterystyczne "wąskie gardło" (1 i 3 warstwa w bloku z jądrem 1x1),
- Przed blokami dodatkowa warstwa konwolucyjna oraz warstwa MaxPool (łącznie 50 warstw)

# Resnet-50

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times 3, 64 \\ 3\times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, 64 \\ 3\times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times 1, 64 \\ 3\times 3, 64 \\ 1\times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times 1, 64 \\ 3\times 3, 64 \\ 1\times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times 1, 64 \\ 3\times 3, 64 \\ 1\times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3\times 3, 128 \\ 3\times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, 128 \\ 3\times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times 1, 128 \\ 3\times 3, 128 \\ 1\times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times 1, 128 \\ 3\times 3, 128 \\ 1\times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1\times 1, 128 \\ 3\times 3, 128 \\ 1\times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3\times 3, 256 \\ 3\times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, 256 \\ 3\times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times 1, 256 \\ 3\times 3, 256 \\ 1\times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1\times 1, 256 \\ 3\times 3, 256 \\ 1\times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1\times 1, 256 \\ 3\times 3, 256 \\ 1\times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3\times 3, 512 \\ 3\times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3\times 3, 512 \\ 3\times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times 1, 512 \\ 3\times 3, 512 \\ 1\times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times 1, 512 \\ 3\times 3, 512 \\ 1\times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1\times 1, 512 \\ 3\times 3, 512 \\ 1\times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		$1.8 \times 10^9$	$3.6 \times 10^9$	$3.8 \times 10^9$	$7.6 \times 10^9$	$11.3 \times 10^9$

architektury sieci rodziny resnet

# Resnet-50

## Przykładowe użycia w problemie reidentyfikacji:

- Peng Wang, Bingliang Jiao, Lu Yang, Yifei Yang, Shizhou Zhang, Wei Wei, Yanning Zhang, *Vehicle Re-Identification in Aerial Imagery: Dataset and Approach*; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 460-469 - Model służący do reidentyfikacji samochodów oparty o resnet50.
- Zuozhuo Dai, Mingqiang Chen, Xiaodong Gu, Siyu Zhu, Ping Tan; *Batch DropBlock Network for Person Re-Identification and Beyond*; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 3691-3701 - dwugałęziowa sieć (Batch DropBlock), gdzie jedną z gałęzi jest konwencjonalny resnet-50

# VGG16

- Należy do rodziny VGGNet.
- Sieć złożona z 13 warstw konwolucyjnych (+maxpooling) oraz 3 warstw w pełni połączonych.
- Po każdym bloku ilość filtrów o rozmiarze 3x3 jest zwiększana dwukrotnie, a wielkość map zmniejszana o połowę.
- Zastosowanie niewielkich filtrów we wszystkich warstwach umożliwia zmniejszenie liczby parametrów w sieci w porównaniu z większymi filtrami. Dodatkowo, zwiększona liczba filtrów w warstwach zwiększa nieliniowość sieci, co z kolei zwiększa jej głębokość w porównaniu z pojedynczym filtrem o większym rozmiarze.

# VGGNet

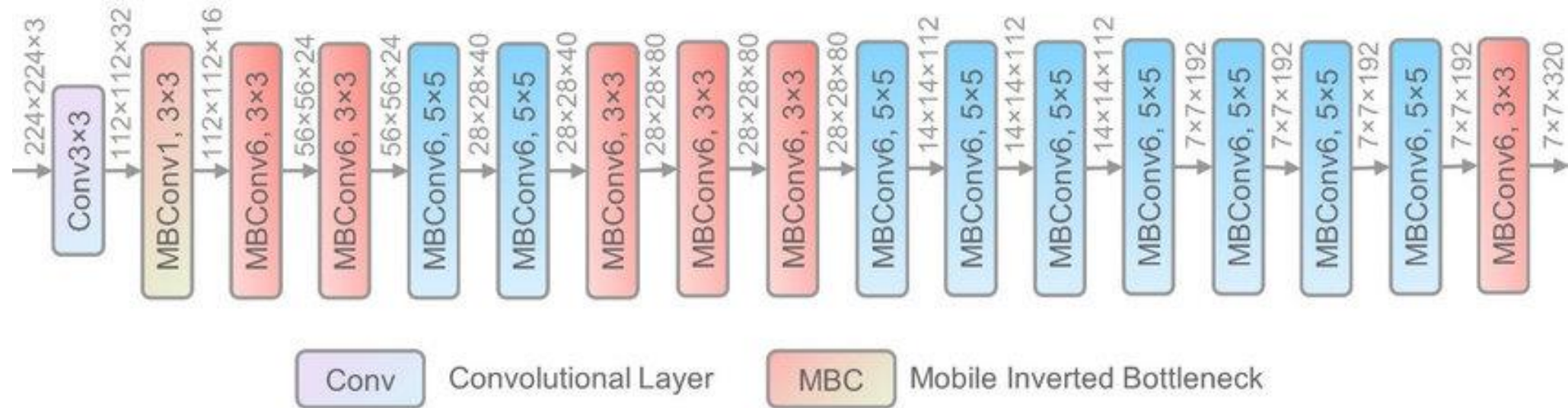
## Architekury sieci

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input ( $224 \times 224$ RGB image)					
conv3-64	conv3-64 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

# EfficientNet B7

- EfficientNet-B7 jest jednym z zaawansowanych modeli z rodziny sieci EfficientNet
- Rodzina ta składa się z skalowalnych, wydajnych konwolucyjnych sieci neuronowych.
- Sieci te są oparte na technice Compound Scaling, która równocześnie skaluje głębokość, szerokość i rozdzielczość sieci.
- EfficientNet-B7 składa się z wielu bloków MBConv, które zawierają konwolucje 1x1, konwolucje z przestrzeni grupowej (DWConv) oraz mechanizm uwagi Squeeze and Excitation (SE).
- W niektórych blokach MBConv wykorzystuje się połączenia typu skip, które łączą wyjście z wcześniejszą warstwą.
- Sieć używa także warstw normalizacji wsadowej (Batch Normalization) oraz funkcji aktywacji Swish.
- Na początku sieci znajduje się warstwa konwolucyjna 3x3 z 64 filtrami.
- Na końcu sieci stosuje się Global Average Pooling (GAP) do agregacji informacji na przestrzeń pojedynczego wektora.
- Warstwa gęsta (Dense) przekształca wektor cech w wartości prawdopodobieństw dla poszczególnych klas.

# Architektura sieci z rodziny EfficientNet



# Bibliografia

- <https://www.getklap.com/blog/what-is-reidentification>,
- <https://datagen.tech/guides/computer-vision/resnet-50/>,
- ZHENG, Zhedong; ZHENG, Liang; YANG, Yi. A discriminatively learned cnn embedding for person reidentification. *ACM transactions on multimedia computing, communications, and applications (TOMM)*, 2017, 14.1: 1-20.
- <https://cv-tricks.com/cnn/understand-resnet-alexnet-vgg-inception/>
- [https://www.researchgate.net/figure/The-network-architecture-of-EfficientNet-It-can-output-a-feature-map-with-deep-semantic\\_fig3\\_349299852/](https://www.researchgate.net/figure/The-network-architecture-of-EfficientNet-It-can-output-a-feature-map-with-deep-semantic_fig3_349299852/)
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. Proceedings of the 36th International Conference on Machine Learning (ICML 2019), 97, 6105–6114.