

lab_4

January 21, 2024

1 Konwolucyjne sieci neuronowe

Dziś spróbujemy stworzyć i wytrenować prostą sieć konwolucyjną do rozpoznawania, co znajduje się na obrazie. Następnie omówimy kwestię identyfikowania obiektów na obrazie, oraz porozmawiamy o wykorzystaniu gotowej już sieci.

1.1 Problem klasyfikacji obrazów

Jak się za to zabrać? Naiwnym podejściem byłaby próba ręcznej specyfikacji pewnych cech (niemowlęta mają duże głowy, szczoteczki są długie, etc.). Szybko jednak stwierdziliśmy, że nawet dla niewielkiego zbioru kategorii jest to tytaniczna praca bez gwarancji sukcesu. Co więcej, istnieje wiele czynników zniekształcających zawartość naszych zdjęć. Obiekty mogą być przedstawiane z różnych ujęć, w różnych warunkach oświetleniowych, w różnej skali, częściowo niewidoczne, ukryte w tle...

Wszystkie wymienione problemy są skutkiem istnienia semantycznej przepaści między tym, jak reprezentowane są nasze dane wejściowe (tablica liczb), a tym, czego w nich szukamy, czyli kategorii i cech: zwierząt, nosów, głów, itp. Zamiast więc próbować samodzielnie napisać funkcję $f(x)$, spróbujemy skorzystać z dobrodziejstw uczenia maszynowego, aby automatycznie skonstruować reprezentację wejścia właściwą dla postawionego sobie zadania (a przynajmniej lepszą od pierwotnej). I tu z pomocą przychodzą nam konwolucyjne sieci neuronowe. Do tego trzeba zrozumieć, czym jest konwolucja (inaczej: splot), a do tego najlepiej nadają się ilustracje, jak to działa.

1.1.1 Konwolucja

Konwolucja (splot) to działanie określone dla dwóch funkcji, dające w wyniku inną, która może być postrzegana jako zmodyfikowana wersja oryginalnych funkcji.

Z naszego punktu widzenia polega to na tym, że mnożymy odpowiadające sobie elementy z dwóch macierzy: obrazu, oraz mniejszej, nazywanej filtrem (lub kernelem). Następnie sumujemy wynik i zapisujemy do macierzy wynikowej na odpowiedniej pozycji. Proces powtarza się aż do momentu przeskanowania całego obrazu. Taki filtr wykrywa, czy coś do niego pasuje w danym miejscu, i z tego wynika zdolność semantycznej generalizacji sieci - uczymy się cech, a wykrywamy je potem w dowolnym miejscu. [Przydatne pojęcia](#)

1.1.2 Stride

Stride to inaczej *krok algorytmu*, albo *przesunięcie*. Określa co ile komórek macierzy następuje aplikacja operatora konwolucji. Jeśli stride wynosi 1 a operator konwolucji ma rozmiar np. 3×3 , to każdy piksel (z wyjątkiem skrajnych narożnych pikseli) będzie uczestniczył w wielu operacjach konwolucji. Jeśli natomiast krok wyniósłby 3, to każdy piksel uczestniczyłby tylko jednokrotnie w tych operacjach. Należy pamiętać, że krok stosujemy zarówno w poziomie, jak i pionie. Najczęściej w obu kierunkach wykorzystuje się ten sam krok.

1.1.3 Padding

Padding to inaczej *wypełnienie* krawędzi obrazu. Określa, w jaki sposób będą traktowane skrajne piksele. Jeśli padding wynosi 0, to skrajne piksele będą uczestniczyły w operacjach konwolucji rzadziej, niż pozostałe piksele (oczywiście jest to również uzależnione od wartości kroku). Aby zniwelować ten efekt, możemy dodać wypełnienie wokół całego obrazu. Te dodatkowe piksele mogą być zerami, albo mogą być również jakimiś uśrednionymi wartościami pikseli sąsiednich. Wypełnienie zerami oznacza de facto obramowanie całego obrazu czarną ramką.

[Więcej na temat wypełnienia.](#)

1.1.4 Pooling

Pooling jest procesem wykorzystywanym do redukcji rozmiaru obrazu. Występują 2 warianty: *max-pooling* oraz *avg-pooling*. Pozwala on usunąć zbędne dane, np. jeżeli filtr wykrywa linie, to istnieje spora szansa, że linie te ciągną się przez sąsiednie piksele, więc nie ma powodu powielać tej informacji. Dzięki temu wprowadzamy pewną inwariancję w wagach sieci i jesteśmy odporni na niewielkie wahania lokalizacji informacji, a skupiamy się na “większym obrazie”.

1.1.5 Sposoby redukcji przeuczenia

- warstwa dropout - wyłączanie losowych neuronów w trakcie uczenia,
- regularyzacja wag - ograniczenie sumy wartości wag,
- metoda wczesnego stopu (early stopping) - zatrzymanie uczenia, jeśli proces uczenia nie poprawia wyników,
- normalizacja paczki (batch normalization) - centrowanie i skalowanie wartości wektorów *w obrębie batcha danych*,
- rozszerzanie danych (data augmentation) - generowanie lekko zaburzonych danych, na podstawie danych treningowych,
- lub... więcej danych.

1.1.6 Budowa sieci CNN do klasyfikacji obrazów

Sieć konwolucyjna składa się zawsze najpierw, zgodnie z nazwą, z części konwolucyjnej, której zadaniem jest wyodrębnienie przydatnych cech z obrazu za pomocą filtrów, warstw poolingowych

etc.

Warstwa konwolucyjna sieci neuronowej składa się z wielu filtrów konwolucyjnych działających równolegle (tj. wykrywających różne cechy). Wagi kerneli, początkowo zainicjalizowane losowo, są dostrajane w procesie uczenia. Wynik działania poszczególnych filtrów jest przepuszczany przez funkcję nieliniową. Mamy tu do czynienia z sytuacją analogiczną jak w MLP: najpierw wykonujemy przekształcenie liniowe, a potem stosujemy funkcję aktywacji. Funkcji aktywacji nie stosuje się jednak po warstwach poolingowych, są to stałe operacje nie podlegające uczeniu.

W celu klasyfikacji obrazu musimy później użyć sieci MLP. Jako że wejściem do sieci MLP jest zawsze wektor, a wyjściem warstwy konwolucyjnej obraz. Musimy zatem obraz przetworzony przez filtry konwolucyjne sprowadzić do formy wektora, tzw. **embedding-u** / **osadzenia**, czyli reprezentacji obrazu jako punktu w pewnej ciągłej przestrzeni. Służy do tego warstwa spłaszczająca (flatten layer), rozwijająca macierze wielkowymiarowe na wektor, np $10 \times 10 \times 3$ na 300×1 .

Część konwolucyjna nazywa się często **backbone**, a część MLP do klasyfikacji **head**. Głowa ma zwykle 1-2 warstwy w pełni połączone, z aktywacją softmax w ostatniej warstwie. Czasem jest nawet po prostu pojedynczą warstwą z softmaxem, bo w dużych sieciach konwolucyjnych ekstrakcja cech jest tak dobra, że taka prosta konstrukcja wystarcza do klasyfikacji embeddingu.

```
[ ]: import torch
import torchvision
import torchvision.transforms as transforms
```

Wybermy rodzaj akceleracji. Współczesne wersje PyTorch wspierają akcelerację nie tylko na kartach Nvidia i AMD, ale również na procesorach Apple z serii M. Obsługa AMD jest realizowana identycznie jak CUDA natomiast MPS (Apple) ma nieco inne API do sprawdzania dostępności i wybierania urządzenia. Zapisujemy wybrane urządzenie do zmiennej `device`, dzięki czemu w dalszych częściach kodu już nie będziemy musieli o tym myśleć.

```
[ ]: if torch.cuda.is_available():
    device = torch.device("cuda:0")
elif torch.backends.mps.is_available():
    device = torch.device("mps")

print(device)
```

`mps`

W pakiecie torchvision mamy funkcje automatycznie pobierające niektóre najbardziej popularne zbiory danych z obrazami.

W tym ćwiczeniu wykorzystamy zbiór FashionMNIST, który zawiera małe (28x28) zdjęcia ubrań w skali szarości. Zbiór ten został stworzony przez Zalando i jest “modowym” odpowiednikiem “cyfrowego” MNIST-a, jest z nim kompatybilny pod względem rozmiarów i charakterystyki danych, ale jest od MNIST-a trudniejszy w klasyfikacji.

Do funkcji ładujących zbiory danych możemy przekazać przekształcenie, które powinno zostać na nim wykonane. Przekształcenia można łączyć przy użyciu `transforms.Compose`. W tym przypadku przekonwertujemy dane z domyślnej reprezentacji PIL.Image na torch-owe tensory.

Pobrany dataset przekazujemy pod kontrolę DataLoader-a, który zajmuje się podawaniem danych w batch-ach podczas treningu.

```
[ ]: transform = transforms.Compose([transforms.ToTensor()])

batch_size = 32

trainset = torchvision.datasets.FashionMNIST(
    root="./data", train=True, download=True, transform=transform
)
trainloader = torch.utils.data.DataLoader(
    trainset, batch_size=batch_size, shuffle=True
)

testset = torchvision.datasets.FashionMNIST(
    root="./data", train=False, download=True, transform=transform
)
testloader = torch.utils.data.DataLoader(
    testset, batch_size=batch_size, shuffle=True
)

classes = (
    "top",
    "Trouser",
    "Pullover",
    "Dress",
    "Coat",
    "Sandal",
    "Shirt",
    "Sneaker",
    "Bag",
    "Ankle boot",
)

print(type(testset[0][0]))
```

```
<class 'torch.Tensor'>
```

Zobaczmy, co jest w naszym zbiorze danych. Poniżej kawałek kodu, który wyświetli nam kilka przykładowych obrazków. Wartości pikseli są znormalizowane do przedziału [0,1].

```
[ ]: import matplotlib.pyplot as plt
import numpy as np

def imshow(img):
    npimg = img.numpy()
    plt.imshow(np.transpose(npimg, (1, 2, 0)))
```

```

plt.axis("off")
plt.show()

dataiter = iter(trainloader)
images, labels = next(dataiter)

def grid_show(images, nrow=8):
    imshow(torchvision.utils.make_grid(images, nrow=nrow))

def print_grid(labels, nrow=8):
    rows = [labels[n : n + nrow] for n in range(0, len(labels), nrow)]
    for r in rows:
        print(" ".join(f"{classes[c]:10s}" for c in r))

grid_show(images)
print_grid(labels)

```



Sneaker	top	Shirt	Ankle boot	Bag	Dress	Sandal
Pullover						
Bag	Dress	Coat	Sandal	Pullover	Sandal	top
Shirt						
Sandal	Sneaker	Ankle boot	Trouser	Bag	Trouser	Sneaker
Shirt						
Bag	Coat	Pullover	Pullover	Ankle boot	Ankle boot	Bag
Ankle boot						

1.2 LeNet

LeNet to bardzo znany, klasyczny model sieci konwolucyjnej.

Warstwy: - obraz - konwolucja, kernel 5×5 , bez paddingu, 6 kanałów (feature maps) - average pooling, kernel 2×2 , stride 2 - konwolucja, kernel 5×5 , bez paddingu, 16 kanałów (feature maps) - average pooling, kernel 2×2 , stride 2 - warstwa w pełni połączona, 120 neuronów na wyjściu - warstwa w pełni połączona, 84 neurony na wyjściu - warstwa w pełni połączona, na wyjściu tyle neuronów, ile jest klas

Zadanie 1 (2 punkty)

Zaimplementuj wyżej opisaną sieć, używając biblioteki PyTorch. Wprowadzimy sobie jednak pewne modyfikacje, żeby było ciekawiej: - w pierwszej warstwie konwolucyjnej użyj 20 kanałów (feature maps) - w drugiej warstwie konwolucyjnej użyj 50 kanałów (feature maps) - w pierwszej warstwie gęstej użyj 300 neuronów - w drugiej warstwie gęstej użyj 100 neuronów

Przydatne elementy z pakietu `torch.nn`: `* Conv2d()` `* AvgPool2d()` `* Linear()`

Z pakietu `torch.nn.functional`: `* relu()`

```
[ ]: import torch.nn as nn
import torch.nn.functional as F

class LeNet(nn.Module):
    def __init__(self):
        super().__init__()

        self.conv1 = nn.Conv2d(1, 20, 5)
        self.pool = nn.AvgPool2d(2, 2)
        self.conv2 = nn.Conv2d(20, 50, 5)
        self.connected1 = nn.Linear(50 * 4 * 4, 300)
        self.connected2 = nn.Linear(300, 100)
        self.connected3 = nn.Linear(100, 10)

    def forward(self, x):
        x = F.relu(self.pool(self.conv1(x)))
        x = F.relu(self.pool(self.conv2(x)))
        x = torch.flatten(x, 1)
        x = F.relu(self.connected1(x))
        x = F.relu(self.connected2(x))
        x = self.connected3(x)
        return x
```

Do treningu użyjemy stochastycznego spadku po gradientie (SGD), a jako funkcję straty Categorical Cross Entropy. W PyTorch-u funkcja ta operuje na indeksach klas (int), a nie na wektorach typu one-hot (jak w Tensorflow).

```
[ ]: import torch.optim as optim

net = LeNet().to(device)

criterion = nn.CrossEntropyLoss()
optimizer = optim.SGD(net.parameters(), lr=0.001, momentum=0.9)
```

Zadanie 2 (1 punkt)

Uzupełnij pętlę uczącą sieć na podstawie jej predykcji. Oblicz (wykonaj krok do przodu) funkcję straty, a następnie przeprowadź propagację wsteczną i wykonaj krok optymalizatora.

```
[ ]: net.train()

for epoch in range(5):
    for data in trainloader:
        inputs, labels = data
        inputs = inputs.to(device)
        labels = labels.to(device)
        outputs = net(inputs)

        loss = criterion(outputs, labels)
        loss.backward()

        optimizer.step()
        optimizer.zero_grad()
```

Zobaczmy na kilku przykładach jak działa wytrenowana sieć.

```
[ ]: dataiter = iter(testloader)
images, labels = next(dataiter)

grid_show(images)
print("Ground Truth")
print_grid(labels)

outputs = net(images.to(device))
_, predicted = torch.max(outputs, 1)

print()
print("Predicted")
print_grid(predicted)
```



Ground Truth

Ankle boot	Trouser	Sandal	Sneaker	Pullover	Coat	Pullover
Pullover						
Sneaker	Sneaker	Bag	Sneaker	Pullover	Shirt	Shirt
Coat						
Pullover	Bag	Dress	Bag	Sneaker	Pullover	Sandal
Coat						
Dress	Ankle boot top		Sandal	Shirt	Coat	top
Pullover						

Predicted

Ankle boot	Trouser	Sandal	Sneaker	Pullover	Shirt	Coat
Pullover						
Ankle boot	Sneaker	Bag	Sneaker	Pullover	top	Pullover
Coat						
Pullover	Bag	Dress	Bag	Sneaker	Pullover	Sandal
Coat						
Dress	Ankle boot top		Sandal	Dress	Coat	top
Pullover						

Obliczmy dokładności (accuracy) dla zbioru danych.

```
[ ]: correct = 0
total = 0
net.eval()
with torch.no_grad():
    for data in testloader:
        images, labels = data
        outputs = net(images.to(device))
        _, predicted = torch.max(outputs.data, 1)
```



```

        total += labels.size(0)
        correct += (predicted == labels.to(device)).sum().item()

print(f"Accuracy of the network on the 10000 test images: {100 * correct //
↪total} %")

```

Accuracy of the network on the 10000 test images: 78 %

Skomentuj wyniki:

- Wynik na poziomie 78% jest dobry, ale nie jest to wynik idealny. Fakt trenowania sieci na 5 epokach może być powodem takiego wyniku. Aby osiągnąć lepszy wynik, można by spróbować zwiększyć liczbę epok.

Znając ogólny wynik klasyfikacji dla zbioru przeanalizujemy dokładniej, z którymi klasami jest największy problem.

Zadanie 3 (1 punkt)

Oblicz dokładność działania sieci (accuracy) dla każdej klasy z osobna. Podczas oceniania skuteczności modelu nie potrzebujemy, aby gradienty się liczyły. Możemy zatem zawrzeć obliczenia w bloku `with torch.no_grad()`:

```

[ ]: from collections import Counter

net.eval()
correct = Counter()
total = Counter()

with torch.no_grad():
    for data in testloader:
        images, labels = data
        outputs = net(images.to(device))
        _, predicted = torch.max(outputs.data, 1)
        for i in zip(labels.to(device).tolist(), predicted.tolist()):
            total.update([i[0]])
            if i[0] == i[1]:
                correct.update([i[0]])

print("Accuracy of the network on the 10000 test images:")
for i in range(10):
    print(f"{classes[i]:10s}: {correct[i]} / {total[i]} = {100 * correct[i] //
↪total[i]} %")

```

Accuracy of the network on the 10000 test images:

```

top       : 895 / 1000 = 89 %
Trouser   : 946 / 1000 = 94 %
Pullover  : 747 / 1000 = 74 %
Dress     : 774 / 1000 = 77 %
Coat      : 724 / 1000 = 72 %
Sandal    : 973 / 1000 = 97 %

```

```
Shirt      : 320 / 1000 = 32 %
Sneaker    : 597 / 1000 = 59 %
Bag        : 937 / 1000 = 93 %
Ankle boot: 946 / 1000 = 94 %
```

Skomentuj wyniki:

- Bardzo łatwo zauważyć, że model ma problem z pewnymi klasami. Najgorzej radzi sobie z koszulką (32%). Ponownie, zwiększenie liczby epok pomogłoby w poprawie wyniku. Gdy przetestowałem model na 15 epokach, to koszulka osiągnęła wynik 54%

1.3 Detekcja obiektów

Problem detekcji polega na nie tylko sklasyfikowaniu obiektów na obrazie, ale również wyznaczeniu jego dokładnego położenia w postaci bounding-box-u. Choć jest to problem odmienny od klasyfikacji obrazów, to w praktyce ściśle z nim powiązany - modele do detekcji obiektów przeważnie do pewnego momentu wyglądają tak samo, jak modele klasyfikacji. Jednak pod koniec sieć jest dzielona na 2 wyjścia: jedno to standardowa klasyfikacja, a drugie to regresor określający pozycję obiektu na obrazie, tzw. bounding box. Najpopularniejszymi przykładami takich sieci są YOLO i Mask R-CNN. Zbiór danych też jest odpowiednio przygotowany do tego zadania i oprócz właściwych zdjęć zawiera również listę bounding-box-ów i ich etykiety.

Zobaczmy jak działa detekcja na przykładzie już wytrenowanej sieci neuronowej. Autorzy skutecznych sieci często udostępniają ich wagi online, dzięki czemu jeżeli mamy doczynienia z analogicznym problemem jak ten, do którego dana sieć była przygotowana możemy z niej skorzystać “prosto z pudełka”.

PyTorch pozwala nam na pobranie wytrenowanych wag dla kilku najpopularniejszych modeli. Sprawdzimy jak z tego skorzystać.

```
[ ]: from torchvision.models import detection
import numpy as np
import cv2
from PIL import Image
import urllib
```

Poniżej znajduje się funkcja pozwalająca wczytać obraz z sieci. Przyda się do testowania działania sieci.

```
[ ]: def url_to_image(url):
    resp = urllib.request.urlopen(url)
    image = np.asarray(bytearray(resp.read()), dtype="uint8")
    image = cv2.imdecode(image, cv2.IMREAD_COLOR)
    return image
```

Model, którym się zajmiemy to Faster R-CNN, który był trenowany na zbiorze COCO. Poniżej znajduje się lista klas (etykiet) dla tego zbioru danych.

```
[ ]: classes = [
    "__background__",
    "person",
```

"bicycle",
"car",
"motorcycle",
"airplane",
"bus",
"train",
"truck",
"boat",
"traffic light",
"fire hydrant",
"street sign",
"stop sign",
"parking meter",
"bench",
"bird",
"cat",
"dog",
"horse",
"sheep",
"cow",
"elephant",
"bear",
"zebra",
"giraffe",
"hat",
"backpack",
"umbrella",
"handbag",
"tie",
"shoe",
"eye glasses",
"suitcase",
"frisbee",
"skis",
"snowboard",
"sports ball",
"kite",
"baseball bat",
"baseball glove",
"skateboard",
"surfboard",
"tennis racket",
"bottle",
"plate",
"wine glass",
"cup",
"fork",

```

"knife",
"spoon",
"bowl",
"banana",
"apple",
"sandwich",
"orange",
"broccoli",
"carrot",
"hot dog",
"pizza",
"donut",
"cake",
"chair",
"couch",
"potted plant",
"bed",
"mirror",
"dining table",
>window",
"desk",
"toilet",
"door",
"tv",
"laptop",
"mouse",
"remote",
"keyboard",
"cell phone",
"microwave",
"oven",
"toaster",
"sink",
"refrigerator",
"blender",
"book",
"clock",
"vase",
"scissors",
"teddy bear",
"hair drier",
"toothbrush",
]

colors = np.random.randint(0, 256, size=(len(classes), 3))

```

Inicjalizacja modelu ResNet50-FPN wytrenowanymi wagami. Inicjalizujemy zarówno sieć backbone jak i RCNN.

```
[ ]: import ssl

ssl._create_default_https_context = ssl._create_unverified_context
```

Bez powyższej komórki nie dało się pobrać modelu.

```
[ ]: model = detection.fasterrcnn_resnet50_fpn(
    weights=detection.FasterRCNN_ResNet50_FPN_Weights.DEFAULT,
    weights_backbone=torchvision.models.ResNet50_Weights.DEFAULT,
    progress=True,
    num_classes=len(classes)
).to(device)
model.eval()
```

Downloading:

```
"https://download.pytorch.org/models/fasterrcnn_resnet50_fpn_coco-258fb6c6.pth"
to /Users/filipdziurdzia/.cache/torch/hub/checkpoints/fasterrcnn_resnet50_fpn_co
co-258fb6c6.pth
100.0%
```

```
[ ]: FasterRCNN(
  (transform): GeneralizedRCNNTransform(
    Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
    Resize(min_size=(800,), max_size=1333, mode='bilinear')
  )
  (backbone): BackboneWithFPN(
    (body): IntermediateLayerGetter(
      (conv1): Conv2d(3, 64, kernel_size=(7, 7), stride=(2, 2), padding=(3, 3),
bias=False)
      (bn1): FrozenBatchNorm2d(64, eps=0.0)
      (relu): ReLU(inplace=True)
      (maxpool): MaxPool2d(kernel_size=3, stride=2, padding=1, dilation=1,
ceil_mode=False)
      (layer1): Sequential(
        (0): Bottleneck(
          (conv1): Conv2d(64, 64, kernel_size=(1, 1), stride=(1, 1), bias=False)
          (bn1): FrozenBatchNorm2d(64, eps=0.0)
          (conv2): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
          (bn2): FrozenBatchNorm2d(64, eps=0.0)
          (conv3): Conv2d(64, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
          (bn3): FrozenBatchNorm2d(256, eps=0.0)
          (relu): ReLU(inplace=True)
          (downsample): Sequential(
            (0): Conv2d(64, 256, kernel_size=(1, 1), stride=(1, 1), bias=False)
            (1): FrozenBatchNorm2d(256, eps=0.0)
          )
        )
      )
    )
  )
```

```

    )
    (1): Bottleneck(
      (conv1): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (bn1): FrozenBatchNorm2d(64, eps=0.0)
      (conv2): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
      (bn2): FrozenBatchNorm2d(64, eps=0.0)
      (conv3): Conv2d(64, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (bn3): FrozenBatchNorm2d(256, eps=0.0)
      (relu): ReLU(inplace=True)
    )
    (2): Bottleneck(
      (conv1): Conv2d(256, 64, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (bn1): FrozenBatchNorm2d(64, eps=0.0)
      (conv2): Conv2d(64, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1), bias=False)
      (bn2): FrozenBatchNorm2d(64, eps=0.0)
      (conv3): Conv2d(64, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (bn3): FrozenBatchNorm2d(256, eps=0.0)
      (relu): ReLU(inplace=True)
    )
  )
  (layer2): Sequential(
    (0): Bottleneck(
      (conv1): Conv2d(256, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (bn1): FrozenBatchNorm2d(128, eps=0.0)
      (conv2): Conv2d(128, 128, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), bias=False)
      (bn2): FrozenBatchNorm2d(128, eps=0.0)
      (conv3): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (bn3): FrozenBatchNorm2d(512, eps=0.0)
      (relu): ReLU(inplace=True)
      (downsample): Sequential(
        (0): Conv2d(256, 512, kernel_size=(1, 1), stride=(2, 2), bias=False)
        (1): FrozenBatchNorm2d(512, eps=0.0)
      )
    )
    (1): Bottleneck(
      (conv1): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (bn1): FrozenBatchNorm2d(128, eps=0.0)

```

```

        (conv2): Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
        (bn2): FrozenBatchNorm2d(128, eps=0.0)
        (conv3): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn3): FrozenBatchNorm2d(512, eps=0.0)
        (relu): ReLU(inplace=True)
    )
    (2): Bottleneck(
        (conv1): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn1): FrozenBatchNorm2d(128, eps=0.0)
        (conv2): Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
        (bn2): FrozenBatchNorm2d(128, eps=0.0)
        (conv3): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn3): FrozenBatchNorm2d(512, eps=0.0)
        (relu): ReLU(inplace=True)
    )
    (3): Bottleneck(
        (conv1): Conv2d(512, 128, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn1): FrozenBatchNorm2d(128, eps=0.0)
        (conv2): Conv2d(128, 128, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
        (bn2): FrozenBatchNorm2d(128, eps=0.0)
        (conv3): Conv2d(128, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn3): FrozenBatchNorm2d(512, eps=0.0)
        (relu): ReLU(inplace=True)
    )
)
(layer3): Sequential(
  (0): Bottleneck(
    (conv1): Conv2d(512, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (bn1): FrozenBatchNorm2d(256, eps=0.0)
    (conv2): Conv2d(256, 256, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), bias=False)
    (bn2): FrozenBatchNorm2d(256, eps=0.0)
    (conv3): Conv2d(256, 1024, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (bn3): FrozenBatchNorm2d(1024, eps=0.0)
    (relu): ReLU(inplace=True)
    (downsample): Sequential(
      (0): Conv2d(512, 1024, kernel_size=(1, 1), stride=(2, 2),

```

```

bias=False)
    (1): FrozenBatchNorm2d(1024, eps=0.0)
    )
    )
    (1): Bottleneck(
        (conv1): Conv2d(1024, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn1): FrozenBatchNorm2d(256, eps=0.0)
        (conv2): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
        (bn2): FrozenBatchNorm2d(256, eps=0.0)
        (conv3): Conv2d(256, 1024, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn3): FrozenBatchNorm2d(1024, eps=0.0)
        (relu): ReLU(inplace=True)
    )
    (2): Bottleneck(
        (conv1): Conv2d(1024, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn1): FrozenBatchNorm2d(256, eps=0.0)
        (conv2): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
        (bn2): FrozenBatchNorm2d(256, eps=0.0)
        (conv3): Conv2d(256, 1024, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn3): FrozenBatchNorm2d(1024, eps=0.0)
        (relu): ReLU(inplace=True)
    )
    (3): Bottleneck(
        (conv1): Conv2d(1024, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn1): FrozenBatchNorm2d(256, eps=0.0)
        (conv2): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
        (bn2): FrozenBatchNorm2d(256, eps=0.0)
        (conv3): Conv2d(256, 1024, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn3): FrozenBatchNorm2d(1024, eps=0.0)
        (relu): ReLU(inplace=True)
    )
    (4): Bottleneck(
        (conv1): Conv2d(1024, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn1): FrozenBatchNorm2d(256, eps=0.0)
        (conv2): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
        (bn2): FrozenBatchNorm2d(256, eps=0.0)

```



```

        (conv3): Conv2d(256, 1024, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn3): FrozenBatchNorm2d(1024, eps=0.0)
        (relu): ReLU(inplace=True)
    )
    (5): Bottleneck(
        (conv1): Conv2d(1024, 256, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn1): FrozenBatchNorm2d(256, eps=0.0)
        (conv2): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
        (bn2): FrozenBatchNorm2d(256, eps=0.0)
        (conv3): Conv2d(256, 1024, kernel_size=(1, 1), stride=(1, 1),
bias=False)
        (bn3): FrozenBatchNorm2d(1024, eps=0.0)
        (relu): ReLU(inplace=True)
    )
)
(layer4): Sequential(
  (0): Bottleneck(
    (conv1): Conv2d(1024, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (bn1): FrozenBatchNorm2d(512, eps=0.0)
    (conv2): Conv2d(512, 512, kernel_size=(3, 3), stride=(2, 2),
padding=(1, 1), bias=False)
    (bn2): FrozenBatchNorm2d(512, eps=0.0)
    (conv3): Conv2d(512, 2048, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (bn3): FrozenBatchNorm2d(2048, eps=0.0)
    (relu): ReLU(inplace=True)
    (downsample): Sequential(
      (0): Conv2d(1024, 2048, kernel_size=(1, 1), stride=(2, 2),
bias=False)
      (1): FrozenBatchNorm2d(2048, eps=0.0)
    )
  )
  (1): Bottleneck(
    (conv1): Conv2d(2048, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (bn1): FrozenBatchNorm2d(512, eps=0.0)
    (conv2): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
    (bn2): FrozenBatchNorm2d(512, eps=0.0)
    (conv3): Conv2d(512, 2048, kernel_size=(1, 1), stride=(1, 1),
bias=False)
    (bn3): FrozenBatchNorm2d(2048, eps=0.0)
    (relu): ReLU(inplace=True)
  )
)

```

```

    )
    (2): Bottleneck(
      (conv1): Conv2d(2048, 512, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (bn1): FrozenBatchNorm2d(512, eps=0.0)
      (conv2): Conv2d(512, 512, kernel_size=(3, 3), stride=(1, 1),
padding=(1, 1), bias=False)
      (bn2): FrozenBatchNorm2d(512, eps=0.0)
      (conv3): Conv2d(512, 2048, kernel_size=(1, 1), stride=(1, 1),
bias=False)
      (bn3): FrozenBatchNorm2d(2048, eps=0.0)
      (relu): ReLU(inplace=True)
    )
  )
)
(fpn): FeaturePyramidNetwork(
  (inner_blocks): ModuleList(
    (0): Conv2dNormActivation(
      (0): Conv2d(256, 256, kernel_size=(1, 1), stride=(1, 1))
    )
    (1): Conv2dNormActivation(
      (0): Conv2d(512, 256, kernel_size=(1, 1), stride=(1, 1))
    )
    (2): Conv2dNormActivation(
      (0): Conv2d(1024, 256, kernel_size=(1, 1), stride=(1, 1))
    )
    (3): Conv2dNormActivation(
      (0): Conv2d(2048, 256, kernel_size=(1, 1), stride=(1, 1))
    )
  )
  (layer_blocks): ModuleList(
    (0-3): 4 x Conv2dNormActivation(
      (0): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1))
    )
  )
  (extra_blocks): LastLevelMaxPool()
)
(rpn): RegionProposalNetwork(
  (anchor_generator): AnchorGenerator()
  (head): RPNHead(
    (conv): Sequential(
      (0): Conv2dNormActivation(
        (0): Conv2d(256, 256, kernel_size=(3, 3), stride=(1, 1), padding=(1,
1))
        (1): ReLU(inplace=True)

```

```

    )
    )
    (cls_logits): Conv2d(256, 3, kernel_size=(1, 1), stride=(1, 1))
    (bbox_pred): Conv2d(256, 12, kernel_size=(1, 1), stride=(1, 1))
    )
    )
    (roi_heads): RoIHeads(
      (box_roi_pool): MultiScaleRoIAlign(featmap_names=['0', '1', '2', '3'],
output_size=(7, 7), sampling_ratio=2)
      (box_head): TwoMLPHead(
        (fc6): Linear(in_features=12544, out_features=1024, bias=True)
        (fc7): Linear(in_features=1024, out_features=1024, bias=True)
      )
      (box_predictor): FastRCNNPredictor(
        (cls_score): Linear(in_features=1024, out_features=91, bias=True)
        (bbox_pred): Linear(in_features=1024, out_features=364, bias=True)
      )
    )
  )
)

```

IPython, z którego korzystamy w Jupyter Notebooku, ma wbudowaną funkcję `display()` do wyświetlania obrazów.

Do pobierania obrazów możemy się posłużyć `wget-em`.

```
[ ]: # Pobieranie obrazka z sieci
# !wget https://upload.wikimedia.org/wikipedia/commons/thumb/7/7a/
↳Toothbrush_x3_20050716_001.jpg/1280px-Toothbrush_x3_20050716_001.jpg
↳--output-document toothbrushes.jpg

--2024-01-21 21:57:26-- https://upload.wikimedia.org/wikipedia/commons/thumb/7/
7a/Toothbrush_x3_20050716_001.jpg/1280px-Toothbrush_x3_20050716_001.jpg
Resolving upload.wikimedia.org (upload.wikimedia.org)... 185.15.59.240
Connecting to upload.wikimedia.org (upload.wikimedia.org)|185.15.59.240|:443...
connected.
HTTP request sent, awaiting response... 200 OK
Length: 338300 (330K) [image/jpeg]
Saving to: 'toothbrushes.jpg'

toothbrushes.jpg  100%[=====>] 330,37K  1,01MB/s   in 0,3s

2024-01-21 21:57:27 (1,01 MB/s) - 'toothbrushes.jpg' saved [338300/338300]
```

```
[ ]: # Wyświetlanie obrazka
image = Image.open("toothbrushes.jpg")
# make sure we have 3-channel RGB, e.g. without transparency
image = image.convert("RGB")
display(image)
```



PyTorch wymaga obrazów w kształcie [channels, height, width] (C, H, W) oraz z wartościami pikseli między 0 a 1. Pillow wczytuje obrazy z kanałami (H, W, C) oraz z wartościami pikseli między 0 a 255. Przed wykorzystaniem sieci neuronowej trzeba zatem: - zamienić obraz na tensor - zmienić kolejność kanałów - podzielić wartości pikseli przez 255

```
[ ]: image_tensor = torch.from_numpy(np.array(image))
image_tensor = image_tensor.permute(2, 0, 1)
image_tensor_int = image_tensor # useful for displaying, dtype = uint8
image_tensor = image_tensor / 255
image_tensor.shape, image_tensor.dtype
```

```
[ ]: (torch.Size([3, 960, 1280]), torch.float32)
```

Zadanie 4 (1 punkt)

Użyj modelu do wykrycia obiektów na obrazie. Następnie wybierz tylko te bounding boxy, dla których mamy wynik powyżej 50%. Wypisz te bounding boxy, ich prawdopodobieństwa (w procentach) oraz nazwy klas.

Następnie wykorzystaj wyniki do zaznaczenia bounding box'a dla każdego wykrytego obiektu na obrazie oraz podpisz wykrytą klasę wraz z prawdopodobieństwem. Możesz tutaj użyć: - [OpenCV](#) - [PyTorch](#) - [Torchvision](#)

```

[ ]: input = image_tensor.unsqueeze(0)
model = model.to("cpu") # doesn't work on Mac mps

with torch.no_grad():
    output = model(input)[0]
mask = output["scores"] > 0.5

boxes = output["boxes"][mask].int().tolist()
labels = np.array(classes)[output["labels"][mask]]
scores = output["scores"][mask]

print(boxes, labels, scores)

for box, label, score in zip(boxes, labels, scores):
    image = cv2.rectangle(
        np.array(image),
        (box[0], box[1]),
        (box[2], box[3]),
        (255, 0, 0),
        2,
    )
    image = cv2.putText(
        image,
        f"{label} {score:.2f}",
        (box[0], box[1]),
        cv2.FONT_HERSHEY_SIMPLEX,
        1,
        (255, 0, 0),
        2,
    )

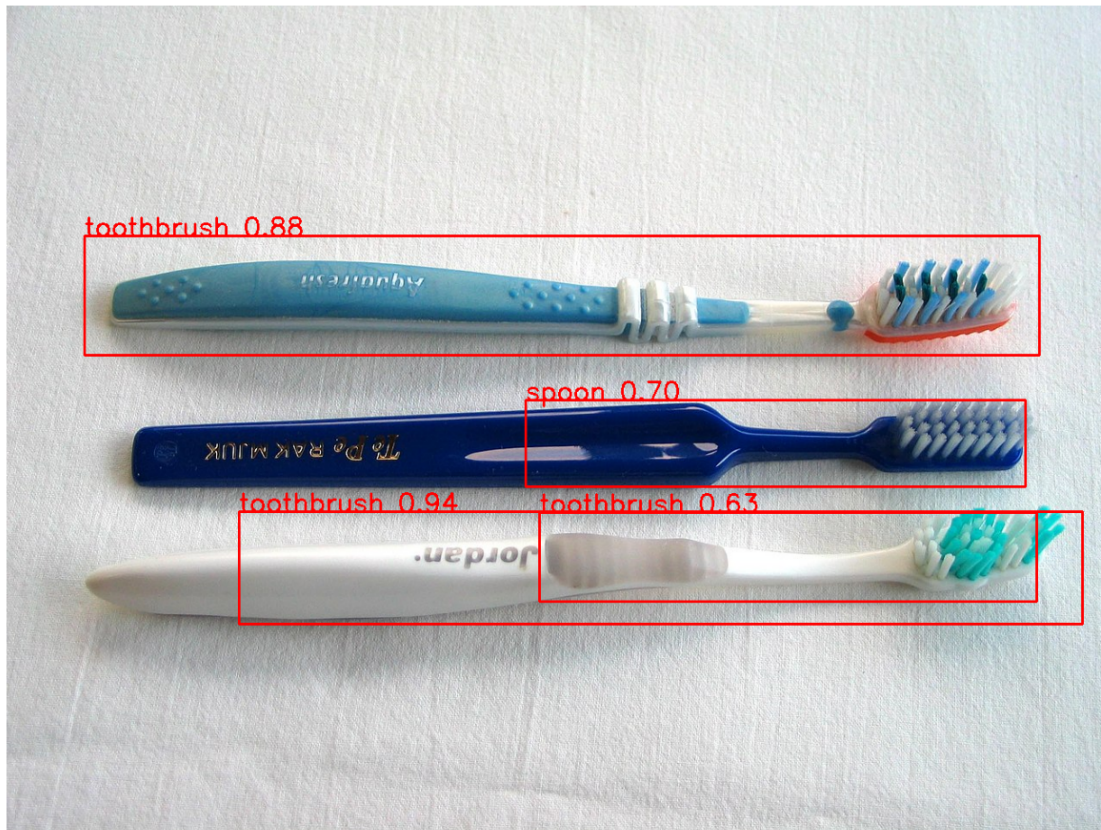
plt.figure(figsize=(16, 9))
plt.axis("off")
plt.imshow(image)
plt.tight_layout()
plt.show()

```

```

[[269, 585, 1246, 715], [90, 266, 1196, 404], [601, 456, 1180, 556], [617, 586,
1193, 689]] ['toothbrush' 'toothbrush' 'spoon' 'toothbrush'] tensor([0.9417,
0.8841, 0.6983, 0.6300])

```

1.4 Fine-tuning i pretrening

Trenowanie głębokich sieci neuronowych do przetwarzania obrazów jest zadaniem wymagającym bardzo dużych zbiorów danych i zasobów obliczeniowych. Często jednak, nie musimy trenować takich sieci od nowa, możemy wykorzystać wytrenowane modele i jedynie dostosowywać je do naszych problemów. Działanie takie nazywa się transfer learning-iem.

Przykładowo: mamy już wytrenowaną sieć na dużym zbiorze danych (pretrening) i chcemy, żeby sieć poradziła sobie z nową klasą obiektów (klasyfikacja), albo lepiej radziła sobie z wybranymi obiektami, które już zna (fine-tuning). Możemy usunąć ostatnią warstwę sieci i na jej miejsce wstawić nową, identyczną, jednak z losowo zainicjalizowanymi wagami, a następnie dotrenować sieć na naszym nowym, bardziej specyficznym zbiorze danych. Przykładowo, jako bazę weźmiemy model wytrenowany na zbiorze ImageNet i będziemy chcieli użyć go do rozpoznawania nowych, nieznanych mu klas, np. ras psów.

Dla przećwiczenia takiego schematu działania wykorzystamy zbiór danych z hotdogami. Będziemy chcieli stwierdzić, czy na obrazku jest hotdog, czy nie. Jako sieci użyjemy modelu ResNet-18, pretrenowanej na zbiorze ImageNet.

```
[ ]: # Download the hotdog dataset
# !wget http://d2l-data.s3-accelerate.amazonaws.com/hotdog.zip
# !unzip -n hotdog.zip
```

```
--2024-01-21 22:04:17-- http://d2l-data.s3-accelerate.amazonaws.com/hotdog.zip
Resolving d2l-data.s3-accelerate.amazonaws.com
(d2l-data.s3-accelerate.amazonaws.com)... 18.244.94.204
Connecting to d2l-data.s3-accelerate.amazonaws.com
(d2l-data.s3-accelerate.amazonaws.com)|18.244.94.204|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: 261292301 (249M) [application/zip]
Saving to: 'hotdog.zip.1'
```

```
hotdog.zip.1          100%[=====>] 249,19M  7,13MB/s    in 44s
```

```
2024-01-21 22:05:01 (5,72 MB/s) - 'hotdog.zip.1' saved [261292301/261292301]
```

Archive: hotdog.zip

Kiedy korzystamy z sieci pretrenowanej na zbiorze ImageNet, zgodnie z [dokumentacją](#) trzeba dokonać standaryzacji naszych obrazów, odejmując średnią i dzieląc przez odchylenie standardowe każdego kanału ze zbioru ImageNet.

All pre-trained models expect input images normalized in the same way, i.e. mini-batches of 3-expected to be at least 224. The images have to be loaded in to a range of $[0, 1]$ and then normalized to $[0.229, 0.225]$. You can use the following transform to normalize:

```
normalize = transforms.Normalize(mean=[0.485, 0.456, 0.406],
                                std=[0.229, 0.224, 0.225])
```

```
[ ]: torch.manual_seed(17)

normalize = transforms.Normalize(
    mean=[0.485, 0.456, 0.406],
    std=[0.229, 0.224, 0.225]
)

train_augs = torchvision.transforms.Compose(
    [
        torchvision.transforms.RandomResizedCrop(224),
        torchvision.transforms.RandomHorizontalFlip(),
        torchvision.transforms.ToTensor(),
        normalize,
    ]
)

test_augs = torchvision.transforms.Compose(
    [
        torchvision.transforms.Resize(256),
        torchvision.transforms.CenterCrop(224),
        torchvision.transforms.ToTensor(),
        normalize,
```

```
]
)
```

```
[ ]: pretrained_net = torchvision.models.resnet18(weights=torchvision.models.
↳ ResNet18_Weights.IMAGENET1K_V1)
```

Downloading: "https://download.pytorch.org/models/resnet18-f37072fd.pth" to
/Users/filipdziurdzia/.cache/torch/hub/checkpoints/resnet18-f37072fd.pth
100.0%

```
[ ]: pretrained_net.fc
```

```
[ ]: Linear(in_features=512, out_features=1000, bias=True)
```

Zadanie 5 (1 punkt)

Dodaj warstwę liniową do naszej fine-tune'owanej sieci oraz zainicjuj ją losowymi wartościami.

```
[ ]: from copy import deepcopy
```

```
finetuned_net = deepcopy(pretrained_net)
finetuned_net.fc = nn.Linear(finetuned_net.fc.in_features, 2)
```

```
[ ]: import time
import copy
```

```
def train_model(
    model, dataloaders, criterion, optimizer, num_epochs=25
):
    since = time.time()

    val_acc_history = []

    best_model_wts = copy.deepcopy(model.state_dict())
    best_acc = 0.0

    for epoch in range(1, num_epochs + 1):
        print("Epoch {}/{}".format(epoch, num_epochs))
        print("-" * 10)

        # Each epoch has a training and validation phase
        for phase in ["train", "val"]:
            if phase == "train":
                model.train() # Set model to training mode
            else:
                model.eval() # Set model to evaluate mode
```



```

running_loss = 0.0
running_corrects = 0

# Iterate over data.
for inputs, labels in dataloaders[phase]:
    inputs = inputs.to(device)
    labels = labels.to(device)

    # zero the parameter gradients
    optimizer.zero_grad()

    # forward
    # track history if only in train
    with torch.set_grad_enabled(phase == "train"):
        # Get model outputs and calculate loss

        outputs = model(inputs)
        loss = criterion(outputs, labels)
        _, preds = torch.max(outputs, 1)

        # backward + optimize only if in training phase
        if phase == "train":
            loss.backward()
            optimizer.step()

    # statistics
    running_loss += loss.item() * inputs.size(0)
    running_corrects += torch.sum(preds == labels.data)

epoch_loss = running_loss / len(dataloaders[phase].dataset)
epoch_acc = running_corrects.float() / len(dataloaders[phase].
↪dataset)

print("{} Loss: {:.4f} Acc: {:.4f}".format(phase, epoch_loss,
↪epoch_acc))

# deep copy the model
if phase == "val" and epoch_acc > best_acc:
    best_acc = epoch_acc
    best_model_wts = copy.deepcopy(model.state_dict())
if phase == "val":
    val_acc_history.append(epoch_acc)

print()

time_elapsed = time.time() - since
print(

```

```

        "Training complete in {:.0f}m {:.0f}s".format(
            time_elapsed // 60, time_elapsed % 60
        )
    )
    print("Best val Acc: {:.4f}".format(best_acc))

    # load best model weights
    model.load_state_dict(best_model_wts)
    return model, val_acc_history

```

```

[ ]: import os

data_dir = "hotdog"
batch_size = 32

model_ft = finetuned_net.to(device)
train_iter = torch.utils.data.DataLoader(
    torchvision.datasets.ImageFolder(
        os.path.join(data_dir, "train"), transform=train_augs
    ),
    batch_size=batch_size,
    shuffle=True,
)
test_iter = torch.utils.data.DataLoader(
    torchvision.datasets.ImageFolder(
        os.path.join(data_dir, "test"), transform=test_augs
    ),
    shuffle=True,
    batch_size=batch_size,
)
loss = nn.CrossEntropyLoss(reduction="none")

```

Zadanie 6 (1 punkt)

Zmodyfikuj tak parametry sieci, aby learning rate dla ostatniej warstwy był 10 razy wyższy niż dla pozostałych.

Trzeba odpowiednio podać pierwszy parametr `torch.optim.SGD` tak, aby zawierał parametry normalne, oraz te z `lr * 10`. Parametry warstw niższych to takie, które mają nazwę inną niż `fc.weight` albo `fc.bias` - może się przydać metoda sieci `named_parameters()`.

```

[ ]: def train_fine_tuning(net, learning_rate, num_epochs=15):

    trainer = torch.optim.SGD(
        [
            {"params": [p for n, p in net.named_parameters() if n not in {"fc.
↪weight", "fc.bias"}]},
            {

```

```

        "params": [p for n, p in net.named_parameters() if n in {"fc.
↪weight", "fc.bias"}],
        "lr": 10 * learning_rate,
    },
],
lr=learning_rate,
)

dataloaders_dict = {"train": train_iter, "val": test_iter}
criterion = nn.CrossEntropyLoss()
model_ft, hist = train_model(
    net, dataloaders_dict, criterion, trainer, num_epochs=num_epochs
)
return model_ft, hist

# your_code

```

```
[ ]: model_ft, hist = train_fine_tuning(model_ft, learning_rate=5e-5)
```

Epoch 1/15

train Loss: 0.7730 Acc: 0.4610

val Loss: 0.6882 Acc: 0.5813

Epoch 2/15

train Loss: 0.6543 Acc: 0.6030

val Loss: 0.5876 Acc: 0.7125

Epoch 3/15

train Loss: 0.5714 Acc: 0.7210

val Loss: 0.5157 Acc: 0.7775

Epoch 4/15

train Loss: 0.5165 Acc: 0.7695

val Loss: 0.4683 Acc: 0.8100

Epoch 5/15

train Loss: 0.4761 Acc: 0.8075

val Loss: 0.4245 Acc: 0.8562

Epoch 6/15

train Loss: 0.4484 Acc: 0.8225

val Loss: 0.3918 Acc: 0.8725

Epoch 7/15

train Loss: 0.4124 Acc: 0.8465

val Loss: 0.3709 Acc: 0.8788

Epoch 8/15

train Loss: 0.3912 Acc: 0.8615

val Loss: 0.3498 Acc: 0.8900

Epoch 9/15

train Loss: 0.3836 Acc: 0.8570

val Loss: 0.3313 Acc: 0.8938

Epoch 10/15

train Loss: 0.3690 Acc: 0.8625

val Loss: 0.3202 Acc: 0.9000

Epoch 11/15

train Loss: 0.3625 Acc: 0.8570

val Loss: 0.3080 Acc: 0.8950

Epoch 12/15

train Loss: 0.3539 Acc: 0.8765

val Loss: 0.3007 Acc: 0.9062

Epoch 13/15

train Loss: 0.3347 Acc: 0.8805

val Loss: 0.2911 Acc: 0.9062

Epoch 14/15

train Loss: 0.3310 Acc: 0.8840

val Loss: 0.2840 Acc: 0.9075

Epoch 15/15

train Loss: 0.3197 Acc: 0.8790

val Loss: 0.2778 Acc: 0.9100

Training complete in 4m 56s

Best val Acc: 0.910000

skomentuj wyniki:

- Dokładność na poziomie 91% to naprawdę dobry wynik. Analizując kolejne epoki, możemy zauważyć, że w ostatnich epokach wzrost dokładności jest bardzo nie wielki, co może sugerować że model jest blisko swojego maksimum.

Przy wyświetlaniu predykcji sieci musimy wykonać operacje odwrotne niż te, które wykonaliśmy, przygotowując obrazy do treningu: - zamienić kolejność kanałów z (C, H, W) na (H, W, C) - zamienić obraz z tensora na tablicę Numpy'a - odwrócić normalizację (mnożymy przez odchylenie standardowe, dodajemy średnią) i upewnić się, że nie wychodzimy poza zakres [0, 1] (wystarczy proste przycięcie wartości)

```
[ ]: def imshow(img, title=None):
    img = img.permute(1, 2, 0).numpy()
    means = np.array([0.485, 0.456, 0.406])
    stds = np.array([0.229, 0.224, 0.225])
    img = stds * img + means
    img = np.clip(img, 0, 1)

    plt.imshow(img)
    if title is not None:
        plt.title(title)

    plt.pause(0.001)
```

```
[ ]: import matplotlib.pyplot as plt
plt.ion()

def visualize_model(model, num_images=6):
    class_names = ["hotdog", "other"]
    model.eval()
    images_so_far = 0
    fig = plt.figure()
    with torch.no_grad():
        for i, (inputs, labels) in enumerate(test_iter):
            inputs = inputs.to(device)
            labels = labels.to(device)

            outputs = model(inputs)
            _, preds = torch.max(outputs, 1)

            for j in range(inputs.size()[0]):
                images_so_far += 1
                ax = plt.subplot(num_images // 2, 2, images_so_far)
                ax.axis('off')
                ax.set_title(f'predicted: {class_names[preds[j]]}')
```

```
imshow(inputs.data[j].cpu())

if images_so_far == num_images:
    return
```

```
[ ]: visualize_model(model_ft)
```

predicted: hotdog



predicted: other



predicted: hotdog



predicted: hotdog



predicted: other



predicted: hotdog



1.5 Zadanie dla chętnych (3 punkty)

W zadaniach dotyczących klasyfikacji obrazu wykorzystywaliśmy prosty zbiór danych i sieć LeNet. Teraz zamień zbiór danych na bardziej skomplikowany, np. [ten](#) lub [ten](#) (lub inny o podobnym poziomie trudności) i zamiast prostej sieci LeNet użyj bardziej złożonej, np. AlexNet, ResNet, MobileNetV2.

[]: