



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

FACULTY OF MECHANICAL ENGINEERING

ÚSTAV AUTOMATIZACE A INFORMATIKY

INSTITUTE OF AUTOMATION AND COMPUTER SCIENCE

**VLIV BAREVNÝCH MODELŮ NA CHOVÁNÍ
KONVOLUČNÍCH NEURONOVÝCH SÍTÍ**

IMPACT OF COLOR MODELS ON PERFORMANCE OF CONVOLUTIONAL NEURAL NETWORKS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. MARTIN ŠIMUNSKÝ

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. PAVEL ŠKRABÁNEK, Ph.D.

BRNO 2020

Zadání diplomové práce

Ústav:	Ústav automatizace a informatiky
Student:	Bc. Martin Šimunský
Studijní program:	Strojní inženýrství
Studijní obor:	Aplikovaná informatika a řízení
Vedoucí práce:	Ing. Pavel Škrabánek, Ph.D.
Akademický rok:	2019/20

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma diplomové práce:

Vliv barevných modelů na chování konvolučních neuronových sítí

Stručná charakteristika problematiky úkolu:

Aktuálně nejlepší systémy počítačového vidění jsou založeny na hlubokých konvolučních neuronových sítích. Tyto systémy často zpracovávají barevné fotografie. Otázkou je, jak barevný model použitý ve fotografiích ovlivňuje chování těchto systémů.

Cíle diplomové práce:

Student připraví rešerši, která bude shrnovat aktuální poznatky o vlivu barevného modelu použitého pro reprezentaci vstupních dat na chování systémů počítačového vidění, které jsou založeny na hlubokých konvolučních neuronových sítích. Student ve spolupráci s vedoucím na základě této rešerše identifikuje oblast vhodnou pro další výzkum. Student navrhne, realizuje a vyhodnotí experimenty relevantní pro identifikovanou oblast výzkumu.

Seznam doporučené literatury:

GONZALEZ, Rafael C. a Richard E. WOODS. Digital image processing. New York, NY: Pearson, [2018]. ISBN 978-0133356724.

ZHAO, Zhong-Qiu, Peng ZHENG, Shou-Tao XU a Xindong WU. Object Detection With Deep Learning: A Review. IEEE Transactions on Neural Networks and Learning Systems [online]. 1-21. DOI: 10.1109/TNNLS.2018.2876865. ISSN 2162-237X. Dostupné z: <https://ieeexplore.ieee.org/document/8627998/>

WANG, Wei a Yujing YANG. Development of convolutional neural network and its application in image classification: a survey. Optical Engineering [online]. 2019, 58(04). DOI: 10.1117/1.OE.58.4.040901. ISSN 0091-3286. Dostupné z: <https://www.spiedigitallibrary.org/journals/optical-engineering/volume-58/issue-04/040901/Development-of-convolutional-neural-network-and-its-application-in-image/10.1117/1.OE.58.4.040901.full>

LATEEF, Fahad a Yassine RUICHEK. Survey on semantic segmentation using deep learning techniques. Neurocomputing [online]. 2019, 338, 321-348 [cit. 2019-09-03]. DOI: 10.1016/j.neucom.2019.02.003. ISSN 09252312. Dostupné z: <https://linkinghub.elsevier.com/retrieve/pii/S092523121930181X>.

GOWDA, Shreyank N. a Chun YUAN. ColorNet: Investigating the Importance of Color Spaces for Image Classification. In: Computer Vision – ACCV 2018 [online]. Cham: Springer International Publishing, 2019, 2019-05-25, s. 581-596. Lecture Notes in Computer Science. DOI: 10.1007/978--030-20870-7_36. ISBN 978-3-030-20869-1. Dostupné z: http://link.springer.com/10.1007/978-3-030-20870-7_36

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2019/20

V Brně, dne

L. S.

doc. Ing. Radomil Matoušek, Ph.D.
ředitel ústavu

doc. Ing. Jaroslav Katolický, Ph.D.
děkan fakulty

Abstrakt

Diplomová práce zkoumá aktuální poznatky týkající se vlivu barevných modelů na chování konvoluční neuronové sítě. Na základě poznatků je proveden experiment na toto téma se šesti barevnými modely RGB, HSV, CIE 1931 XYZ, CIE 1976 L*a*b*, YIQ a YCbCr a hlubokou konvoluční neuronovou sítí ResNet-101. Barevný model RGB dosáhl v tomto experimentu nejvyšší přesnosti klasifikace, model HSV nejnižší.

Abstract

Current knowledge about impact of colour models on performance of convolutional neural network is investigated in the first part of this thesis. The experiment based on obtained knowledge is conducted in the second part. Six colour models HSV, CIE 1931 XYZ, CIE 1976 L*a*b*, YIQ and YCbCr and deep convolutional neural network ResNet-101 are used. RGB colour model achieved the highest classification accuracy, whereas HSV color model has the lowest accuracy in this experiment.

Klíčová slova

barevné modely, porovnání, vliv, klasifikace objektů, konvoluční neuronová síť, počítačové vidění, hluboké učení, rgb, hsv, xyz, lab, yiq, ycbr

Keywords

color models, colour models, comparison, image classification, object classification, impact, convolutional neural network, computer vision, deep learning, rgb, hsv, xyz, lab, yiq, ycbr

Citace

ŠIMUNSKÝ, Martin. *Vliv barevných modelů na chování konvolučních neuronových sítí*. Brno, 2020. Diplomová práce. Vysoké učení technické v Brně, Fakulta strojního inženýrství. Vedoucí práce Ing. Pavel Škrabánek, Ph.D.

Vliv barevných modelů na chování konvolučních neuronových sítí

Prohlášení

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně pod vedením Ing. Pavla Škrabánka, Ph.D. Literaturu, kterou jsem při vypracování použil, důsledně uvádím na konci diplomové práce.

.....
Martin Šimunský
26. června 2020

Poděkování

Tímto bych rád poděkoval svému vedoucímu Ing. Pavlu Škrabánkovi, Ph.D. za vstřícný, ochotný a trpělivý přístup, poskytnuté konzultace a cenné rady během zpracovávání této práce. Veliké díky patří také mé rodině, blízkým a všem, kteří mě jakkoliv podporovali nejen při psaní práce, ale i během celého studia.

Obsah

1	Úvod	10
2	Teoretická část	12
2.1	Konvoluční neuronové sítě	12
2.1.1	Historie hlubokého učení	12
2.1.2	Využití hlubokého učení v počítačovém vidění	14
2.1.3	Základní stavební prvky CNN	15
2.1.4	Moderní architektury CNN	18
2.1.5	Transfer learning, matice záměn a přesnost trénování	21
2.2	Barevné modely	22
2.2.1	RGB barevný model	23
2.2.2	HSV barevný model	24
2.2.3	XYZ barevný prostor	26
2.2.4	LAB barevný prostor	28
2.2.5	YIQ barevný model	30
2.2.6	YCbCr barevný model	30
2.3	Poznatky o vlivu barevných modelů na chování CNN	31
3	Praktická část	34
3.1	Popis experimentu	34
3.2	Datová sada	35
3.3	Inicializační experiment	36
3.4	Hlavní experiment	38
3.5	Samostatné vyhodnocení modelu HSV	42
4	Závěr	46
	Literatura	47

Seznam obrázků

2.1	Přibližný počet výsledků vyhledávání za daný rok podle klíčového slova <i>convolutional neural network</i> . Vyhledávání bylo provedeno na dvou různých webových vyhledávačích. Porovnání vyhledávače Google Scholar a vyhledávače Primo Vysokého učení technického v Brně.	13
2.2	Grafické znázornění XOR	13
2.3	Struktura jednoduché CNN. Převzato z [6].	15
2.4	Znázornění operace konvoluce – velikost konvolučního jádra je 3×3 , krok je roven 1 a <i>padding</i> je roven 0. Je vidět zmenšení vstupní vrstvy, což je u konvoluce nežádoucí.	16
2.5	Znázornění operace konvoluce – velikost konvolučního jádra je 3×3 , krok je roven 1 a <i>padding</i> je roven 1	17
2.6	Znázornění operace sdružování dle maxima přes oblast o velikosti 2×2 , krok je roven 2 a <i>padding</i> je roven 0	18
2.7	Schématické znázornění reziduálního <i>bottleneck</i> bloku využívaného v architektuře ResNet-101. Vstupní vrstva má hloubku 256 (d na vstupu a výstupu znamená hloubku vrstvy). V levém případě první konvoluce 1×1 redukuje hloubku vrstvy na 64. Po každé konvoluci následuje <i>batch</i> normalizace. Třetí konvoluce zvýší hloubku vrstvy opět na rozměr 256, aby odpovídala vstupní vrstvě a mohlo dojít k součtu obou větví po prvcích, což představuje symbol kolečka. Vpravo je hloubka výstupní vrstvy navyšována, a tak je zde navíc konvoluce 1×1 s <i>batch</i> normalizací.	19
2.8	Schéma prvního <i>dense</i> bloku architektury DenseNet-201, d je hloubka vrstvy. Tempo růstu (<i>growth rate</i>) je v tomto případě $k = 32$. Celkem má tento <i>dense</i> blok 6 větví, které jsou složeny ze dvou <i>batch</i> normalizací, dvou ReLU, konvoluce 1×1 a 3×3 . Na obrázku jsou vyznačeny pouze dvě větve z důvodu úspory místa, zbylé čtyři je možno si domyslet namísto tečkované šipky.	20
2.9	Schématické znázornění prvního <i>inception</i> modulu architektury Inception v1 (GoogLeNet). Na konci modulu dojde ke zřetězení příznakových map, takže hloubka na výstupu tohoto modulu je $64 + 128 + 32 + 32 = 256$. V pravé větvi je provedeno sdružování dle maxima o velikosti 3×3 a krokem 1.	21
2.10	Matrice záměn vzatá z vyhodnocení experimentu pro barevný model RGB	22

2.11	Znázornění barevného modelu RGB. V počátku souřadnicového systému $[0, 0, 0]$ je černá barva. Bílá barva se nachází na souřadnici $[255, 255, 255]$. Spojnice (úhlopříčka) těchto dvou bodů představuje stupně šedi. Na jednotlivých osách R, G, B jsou základní barvy, doplňkové barvy se nachází na zbylých třech vrcholech a vzniknou vždy smícháním dvou základních barev: azurová z modré a zelené, purpurová z červené a modré a žlutá z červené a zelené. Vytvořeno pomocí aplikace MATLAB.	24
2.12	Znázornění barevného modelu HSV. Odstín H je dán úhlem průvodiče v základně kuželu, sytost S a relativní světlost V je v rozmezí $\langle 0, 1 \rangle$	26
2.13	Chromatický diagram CIE xy s gamutem sRGB a standardním osvětlením D65 (bílý bod). Osy představují chromatické souřadnice x a y , hraniční křivka tvaru podkovy se nazývá spektrální locus a ukazuje vlnové délky čistých spektrálních barev. Čára spojující koncové body nese jméno purpurová linie a jsou na ní zobrazeny nespektrální barvy. Převzato z [25].	28
3.1	Architektury CNN nabízené MATLABem, na vodorovné ose je čas potřebný k tréninku dané architektury vztažený k nejrychlejší architektuře. Přesný čas závisí na použitém hardwaru a velikosti <i>mini-batch</i> . Svislou osu představuje přesnost klasifikace v procentech a velikosti bublin znázorňují velikost architektury v MB. Převzato z [28].	35
3.2	Porovnání jednotlivých barevných modelů z hlediska přesnosti testování	40
3.3	Rozkmitání trénovací křivky v posledních iteracích má za následek snížení testovací přesnosti	40
3.4	Porovnání dvou tvarově a barevně podobných obrázků ze dvou různých kategorií datové sady Food-101. Převzato z [3].	42
3.5	Trénovací křivka u inicializačního experimentu barevného modelu HSV. Celkem 20 000 iterací, tzn. 100 epoch.	43
3.6	Porovnání barevných modelů RGB a HSV obrázku z kategorie sushi datové sady Food-101. U modelu HSV lze pozorovat diskontinuitu barev. Převzato z [3].	44
3.7	Porovnání jednotlivých kanálů barevného modelu HSV obrázku z kategorie sushi datové sady Food-101. Převzato z [3].	44
3.8	Znázornění diskontinuity transformace z barevného modelu RGB do modelu HSV. Vlevo nahoře jsou zobrazeny dva pixely červené vedle sebe, jejichž rozdíl v barvě je nerozeznatelný, ale transformace do modelu HSV se projeví v kanále H ostrým přechodem z hodnoty 0 (černá) do hodnoty 1 (bílá), což je znázorněno vpravo nahoře. Dolní obrázek vlevo podobně ukazuje dva téměř totožné pixely černé v modelu RGB a vpravo pak jejich reprezentaci v kanálu S modelu HSV – opět lze vidět ostrý přechod při přidání jednoho bitu do jakékoli složky modelu RGB.	45

Seznam tabulek

3.1	Nastavení hyperparametrů pro inicializační experiment	37
3.2	Shrnutí výsledků inicializačního experimentu	37
3.3	Přesnosti testování pro různé nastavení míry učení	38
3.4	Nastavení hyperparametrů pro hlavní experiment	38
3.5	Shrnutí výsledků hlavního experimentu	39
3.6	Průměry falešně negativních odhadů pro všechny kategorie všech barevných modelů. Červeně jsou označeny nejhorší výsledky, zeleně nejlepší.	41
3.7	Mediány falešně negativních odhadů pro všechny kategorie všech barevných modelů. Červeně jsou označeny nejhorší výsledky, zeleně nejlepší.	41

Kapitola 1

Úvod

Konvoluční neuronová síť je pojem, se kterým je možno se v posledních pěti letech setkávat stále častěji. Téma to není nové. Pomyslný stavební kámen hlubokého učení, do něhož konvoluční neuronové sítě patří, položili již v roce 1943 dvojice amerických vědců Warren McCulloch a Walter Pitts. Nicméně až díky dnešnímu dostatečnému výkonu počítačů se konvoluční neuronové sítě dokázaly proslavit. Roku 2012 vyhrály soutěž v rozpoznávání obrázků ILSVRC [19], v roce 2015 překonaly v přesnosti klasifikace člověka [14] a roku 2016 porazily mistra v asijské deskové hře Go [29]. Uplatňují se ale i mimo zábavní průmysl. Počítače díky nim dokáží zpracovat informace ze zvukového záznamu nebo řeči, obrazu nebo videa, a tak například jsou schopné detekovat nádory z lékařských snímků. Je možno rozpoznat na fotkách obličej i jejich mimiku, zjistit přítomnost požáru či zloděje z videozáznamu nebo vyhodnotit pravost bankovky. Velké využití nachází konvoluční neuronové sítě i v automobilovém průmyslu, například při rozpoznávání dopravního značení nebo chodců ve vozovce. Důvody, které staví konvoluční neuronové sítě do popředí vědeckého zájmu jsou hlavně vysoká přesnost rozpoznávání, která souvisí s narůstajícím množstvím trénovacích dat a dále jejich jednoduchá implementace.

Pro trénování konvolučních neuronových sítí jsou použita obrazová data, která jsou primárně reprezentována v barevném modelu RGB. Tato data je nicméně možné reprezentovat i v jiných barevných modelech a použít je pro úlohu klasifikace objektů. Cílem této diplomové práce je prozkoumat aktuální poznatky vlivu některých barevných modelů na chování konvoluční neuronové sítě. Autor se domnívá, že toto téma je „relativně“ nové a systematické shrnutí zatím chybí, a proto je záhodno zjistit současný stav poznání a na základě těchto poznatků zhotovit experiment o tom, jaký mají některé barevné modely vliv na chování konvolučních neuronových sítí.

Práce je rozčleněna následovně. První kapitola připadá tomuto úvodu, druhá je věnována teorii konvolučních neuronových sítí (dále v textu je pro ně používaná zkratka CNN), barevných modelů a na závěr kapitoly je shrnutý současný stav

poznání vlivu barevných modelů na chování konvolučních neuronových sítí. Třetí, praktická, kapitola se týká samotného experimentu. V poslední, čtvrté, kapitole jsou shrnuty výsledky experimentu a je vyvozeno doporučení pro další výzkumnou činnost, jež vychází z poznatků nabytých v průběhu tvorby této závěrečné práce.

Kapitola 2

Teoretická část

Teoretická část nejprve rozvíjí téma konvolučních neuronových sítí (dále jen CNN). Je pojednáno o historii, CNN jsou zařazeny do kontextu umělé inteligence a princip fungování jednoduché CNN je popsán. Následuje ukázka některých známých architektur CNN. Na závěr tohoto celku o CNN je zmíněna metoda využití předtrénované CNN zvaná *transfer learning*, je pojednáno o matici záměn a vyhodnocení přesnosti testování.

Další podkapitola shrnuje teorii barevných modelů, s nimiž je pracováno v praktické části. Jedná se o barevné modely RGB, HSV, CIE 1931 XYZ, CIE 1976 L*a*b*, YIQ a YCbCr. Na závěr této kapitoly jsou prohledány a shrnuty dostupné informační zdroje na téma vlivu barevných modelů na chování CNN.

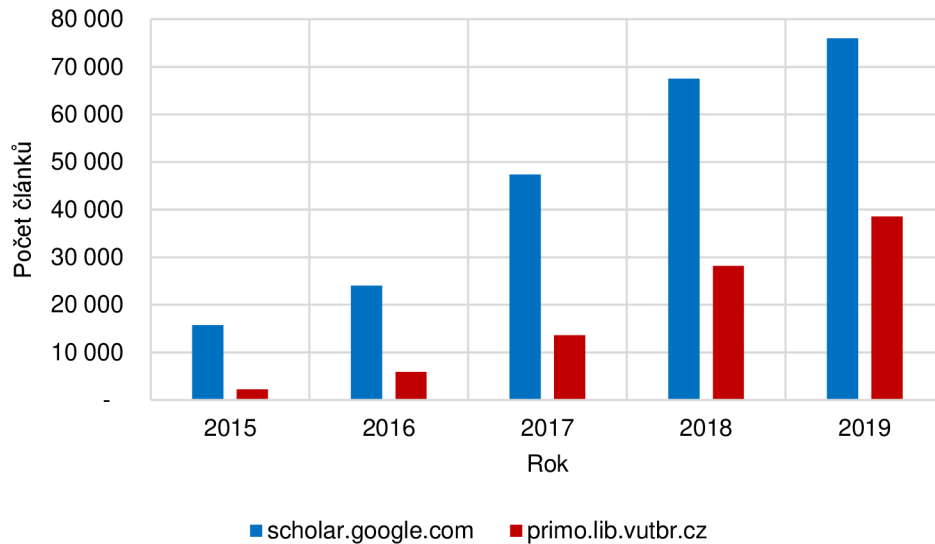
2.1 Konvoluční neuronové sítě

Obliba a zájem o CNN v posledních pěti letech rapidně narůstá. Nasvědčuje tomu rostoucí počet publikovaných článků. Na [Obr. 2.1](#) lze vidět srovnání vyhledávače Google Scholar a Primo, což je vyhledávač Vysokého učení technického v Brně.

2.1.1 Historie hlubokého učení

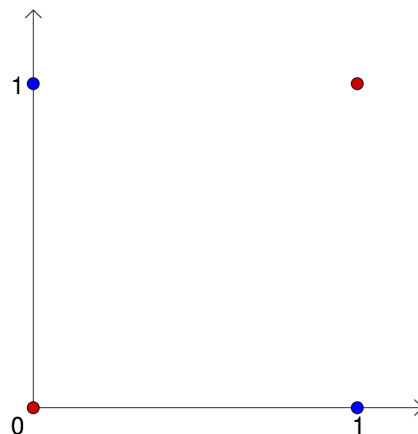
Tato podkapitola čerpá převážně z knihy *Deep Learning* [12]. Historii hlubokého učení (deep learning) je možno započít v dobách druhé světové války. V roce 1943 dvojice amerických vědců Warren McCulloch a Walter Pitts ve Spojených státech sestrojili první jednoduchý matematický model neuronu. Rok 1957 se pojí se jménem Frank Rosenblatt, jenž se zasadil o perceptron – první model umělé neuronové sítě. Nicméně ve stejném roce, kdy člověk poprvé chodil po Měsíci (1969), byl publikován rukopis Marvina Minského a Seymoura Paperta v knize *Perceptrons: An*

Introduction to Computational Geometry, kteří svým argumentem o tom, že Rosenblattův jednovrstvý perceptron nedokáže řešit lineárně separovatelný problém XOR, podpořil všeobecný akademický pesimismus, jenž v té době vládl. Výzkum neuronových sítí byl zpomalen, jelikož jeho podpora klesla.



Obr. 2.1 Přibližný počet výsledků vyhledávání za daný rok podle klíčového slova *convolutional neural network*. Vyhledávání bylo provedeno na dvou různých webových vyhledávačích. Porovnání vyhledávače Google Scholar a vyhledávače Primo Vysokého učení technického v Brně.

Problém XOR si je možno představit tak, jak je znázorněno na Obr. 2.2. Čtyři body jsou v rovině uspořádané do čtverce. Dva modré body se nachází přes jednu uhlopříčku a dva červené přes druhou uhlopříčku. Cílem je rozdělit tyto body jedinou přímkou tak, aby na jedné straně přímky byly jen modré body a na druhé straně červené body. Vidno, že samotnou přímkou tento problém nelze vyřešit. Daná úloha není lineárně separovatelná.



Obr. 2.2 Grafické znázornění XOR

Roku 1979 představil Kunihiro Fukushima základ pro současné CNN zvaný *neocognitron*, který dokázal zpracovat obrázky. Obrození však nastalo až ve druhé polovině 80. let, kdy byl zpopularizován (nikoli objeven) algoritmus zpětného šíření chyby (backpropagation). Komerčně byly CNN používány dokonce již od roku 1993 pro optické rozpoznávání znaků (OCR) a pro rozpoznávání ručně psaného textu. Zakomponovány byly ve velkých bankách, kde sloužily ke čtení šeků. Na konci devadesátých let CNN četly přes 10 % všech šeků v USA [20].

Od roku 2012 dosahují CNN prvních příček v soutěži ImageNet Large Scale Visual Recognition Challenge (ILSVRC). ImageNet je obrovská obrazová datová sada určená hlavně pro výzkum v úloze klasifikace obrázků. Prvním průkopníkem byla v tomto roce architektura AlexNet, která vyhrála první místo [19].

Hluboké učení je využíváno korporacemi jako jsou Google, Microsoft, Facebook, Netflix či Apple. Jako zajímavost stojí za to zmínit společnost DeepMind, která se spojila se společností Google v roce 2014. Tato společnost vytvořila počítačový program AlphaGo, který jako první dokázal porazit druhého nejlepšího hráče světa deskové hry Go Leea Sedola v roce 2016 [29]. Společnost OpenAI, se kterou je spojeno jméno Elon Musk, vytvořila projekt OpenAI Five, což je tým pěti neuronových sítí, přičemž každá hraje za jednu postavu ve hře Dota 2. V dubnu 2019 porazil tento tým světové šampióny v turnaji OG [24].

2.1.2 Využití hlubokého učení v počítačovém vidění

CNN je třídou taxonomie hlubokého učení. Hluboké učení spadá do širší rodiny strojového učení (machine learning). Nakonec strojové učení je jedním z přístupů umělé inteligence (artificial intelligence, AI). Umělá inteligence tedy pomyslně stojí na vrcholku hierarchie.

Další oblastí je počítačové vidění (computer vision), do něhož spadá strojové vidění. To se zabývá získáváním informací z obrazu a jeho zpracováním. Do zpracování obrazu se řadí i barevné modely. Počítačové vidění v některých aplikacích využívá umělou inteligenci. Mezi aplikace počítačového vidění patří [4]:

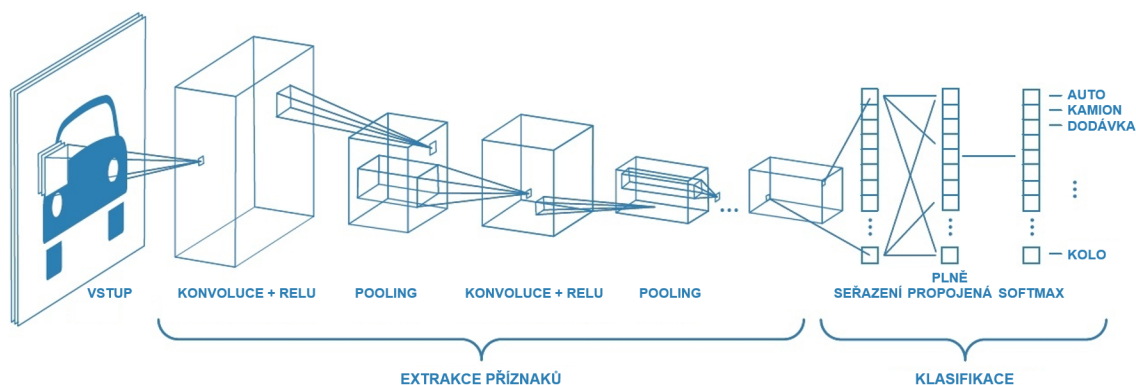
- Klasifikace objektu: do jaké kategorie spadá objekt v obrázku?
- Identifikace objektu: jaký je to typ předložené kategorie v obrázku?
- Ověření objektu: je ten objekt v obrázku přítomen?
- Detekce objektu: kde se objekt v obrázku nachází?
- Segmentace objektu: které pixely přísluší danému objektu v obrázku?
- Rozpoznání objektu: co jsou v obrázku za objekty a kde jsou?

CNN lze pro výše zmíněné aplikace počítačového vidění použít. Na vstupu CNN možno poskytnout např. zvuková data, primárně jsou ale využívána data obrazová, která jsou nejčastěji reprezentována v barevném modelu RGB. Pokud je poskytnut na vstupu CNN obrázek v jiném modelu než RGB, je možno dosáhnout rozdílných výsledků. Tento aspekt je v práci hlouběji zkoumán. [34]

2.1.3 Základní stavební prvky CNN

CNN jsou jednou z tříd taxonomie hlubokého učení. Je vhodné tedy začít teorií o samotném hlubokém učení. Dříve počítače dokázaly řešit (a stále řeší) problémy, které byly náročné pro člověka, ale jednoduché pro počítače (vyžadovaly spoustu výpočtů, které počítače zvládaly za zlomek času lidské snahy). Opravdová výzva přišla v momentě, kdy měly počítače řešit problémy, které jsou pro lidstvo zcela intuitivní, ale pro počítač náročné. Jedná se například o rozpoznávání tváře z fotografie. Řešení nabízí hluboké učení. To je založeno na psychologické koncepci konekcionismu. Zatímco jeden biologický neuron není inteligentní, velké množství vzájemně propojených neuronů vykazuje inteligentní chování. Velikost neuronové sítě je tedy kritickým parametrem a hluboké učení má vysoké nároky na výpočetní výkon. [12]

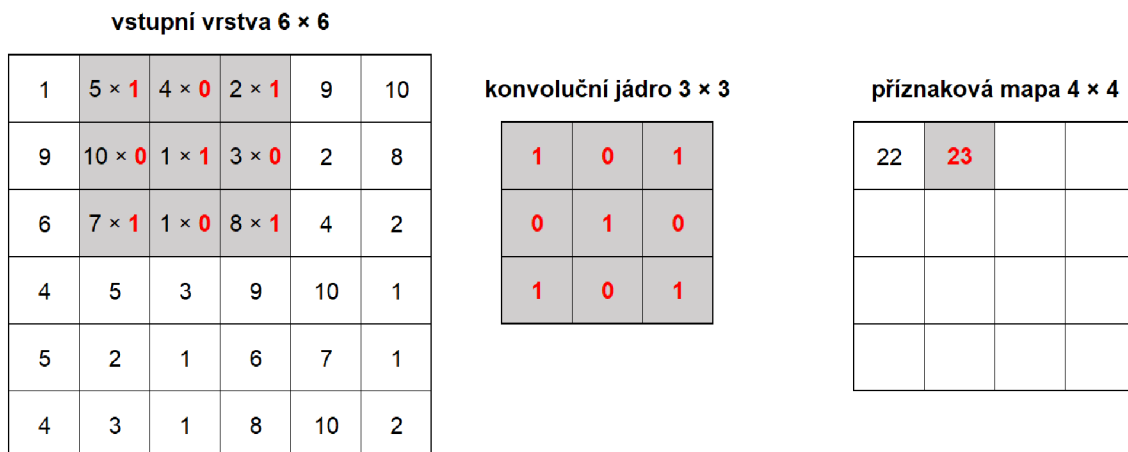
Mezi metody strojového učení patří učení s učitelem a bez něj. U metody učení bez učitele nejsou klasifikátoru poskytnuta trénovací data. U metody učení s učitelem jsou naopak data klasifikátoru poskytnuta. V kontextu této práce je řeč právě o metodě učení s učitelem. Trénovací data v podobě obrázků jsou poskytnuta CNN na vstupu. Obrázek prochází sérií konvolučních vrstev, aktivačních funkcí a sdružovacích vrstev (pooling). Část CNN složená z těchto vrstev je známa jako extrakce příznaků (Obr. 2.3). První vrstvy extrahují jednoduché tvary, hrany a linie. Hlubší vrstvy extrahují komplexnější tvary a kontury, které jsou z jednoduchých tvarů složeny.



Obr. 2.3 Struktura jednoduché CNN. Převzato z [6].

Další část CNN (klasifikační) je složena z neuronů, které jsou plně propojeny. Tato část převezme extrahované příznaky a provádí klasifikaci. Výstupem CNN je softmax funkce – pravděpodobnost, s jakou objekt v obrázku na vstupu spadá do předem stanovených kategorií. Počet neuronů v poslední vrstvě tedy odpovídá počtu kategorií klasifikační úlohy.

Operace konvoluce je provedena posouváním konvolučního jádra po vstupním obrázku. Jádro má pevně dané hodnoty (váhy) a jeho rozměry (šířka a výška) mohou nabývat velikosti například 3×3 . Při každém posunutí dojde nejdříve k vynásobení všech prvků na sobě ležících matic (násobení po prvcích). Výsledných devět součinů (podle počtu prvků konvolučního jádra) je sečteno a dospěje se tedy k jednomu výsledku. Toto číslo je zapsáno do nové matice zvané příznaková mapa (feature map). Poté je posunuto konvoluční jádro o daný krok a proces je zopakován, dokud není takto vyřešena celá vstupní vrstva. Operace konvoluce je znázorněna na **Obr. 2.4**.



Obr. 2.4 Znázornění operace konvoluce – velikost konvolučního jádra je 3×3 , krok je roven 1 a *padding* je roven 0. Je vidět zmenšení vstupní vrstvy, což je u konvoluce nežádoucí.

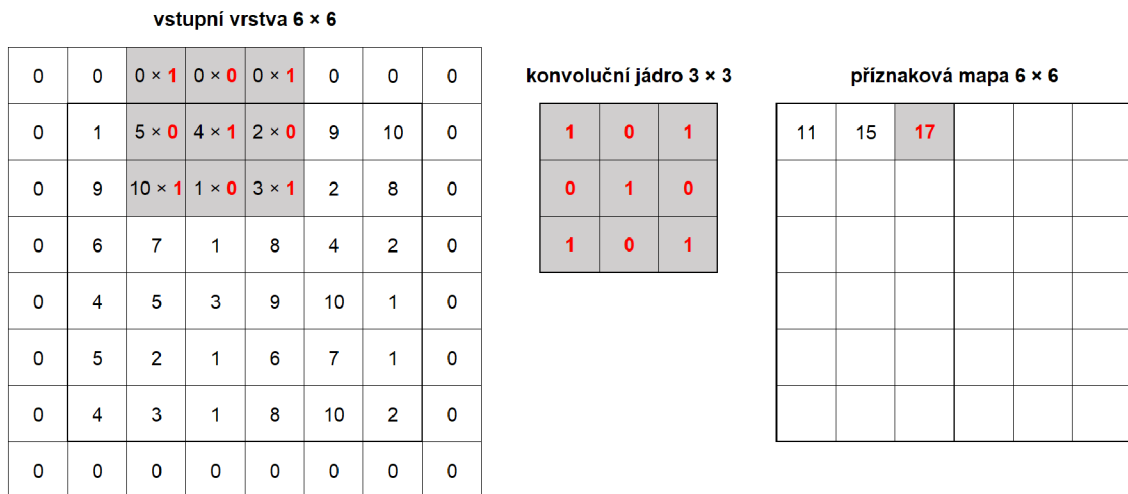
Je potřeba dodat, že toto je pouze jedna operace konvoluce a jedna příznaková mapa. V jedné vrstvě je vždy provedeno více konvolucí s rozdílnými konvolučními jádry lišícími se svými hodnotami vah. Vzniká tedy více příznakových map, čímž narůstá hloubka druhé vrstvy.

Příkladem¹ budiž první konvoluční vrstva architektury ResNet-101. Tato vrstva je čtyřrozměrné pole. První dva rozměry 7×7 jsou rozměry konvolučního jádra – šířka a výška. Třetí rozměr nabývá hodnot 1 až 3 a jedná se o kanál vstupního barevného obrazu, pro něhož je konvoluce provedena. Čtvrtý rozměr je v rozmezí 1 až 64 a jedná se o hloubku této vrstvy. Výstupem je v tomto případě 64 příznakových map – trojrozměrné pole. Každá příznaková mapa má totiž jen jeden kanál, jelikož

¹ Zde je již řeč o 3D konvoluci. Vstupní obraz není jednokanálový, ale tříkanálový. Příklad je vzatý z MATLABu, kde je architektura ResNet-101 zobrazena pomocí aplikace Deep Network Designer.

operace konvoluce v případě barevného obrazu spočívá v sečtení každého součinu všech tří kanálů.

Obvykle je na vstupní vrstvě využíván *padding*. Jedná se o rámeček nulových hodnot. Tímto se docílí toho, že příznaková mapa bude mít stejné rozměry (šířku a výšku) jako vstupní vrstva. *Paddingu* je u konvolučních neuronových sítí využíváno, protože architektura CNN je složena hlavně z konvolučních vrstev. Bez *paddingu* by docházelo ke zmenšování rozměrů, což není žádané. Schéma operace konvoluce s *paddingem* 1 lze vidět na Obr. 2.5.



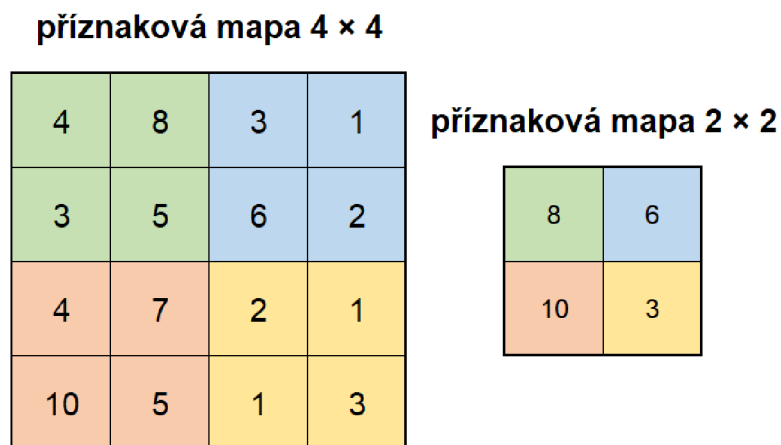
Obr. 2.5 Znárodnění operace konvoluce – velikost konvolučního jádra je 3×3 , krok je roven 1 a *padding* je roven 1

Výsledek, tedy každý prvek příznakové mapy, je poslán přes aktivační funkci. Bez této funkce by se jednalo pouze o lineární operace, a tudíž by neuronová síť měla velmi omezené možnosti. Obvykle se jedná o nelineární funkci ReLU (rectified linear unit). Pokud je vstup do této funkce záporný, výstupem je hodnota nula. V případě vstupu kladné hodnoty je výstup roven hodnotě vstupu.

Po konvoluci a ReLU je provedeno tzv. sdružování (pooling), které snižuje první dva rozměry trojrozměrné příznakové mapy – jejich šířku a výšku, hloubka je zachována. Zde je to žádané, jelikož touto metodou je snížen počet parametrů, což snižuje tréninkový čas a tento krok předchází přeučení CNN (overfitting). Nejběžnějším sdružováním je sdružování dle maxima (max pooling), které z daného počtu sousedících pixelů vybere maximální hodnotu a uloží ji do nové příznakové mapy, jak zobrazuje obrázek Obr. 2.6. U sdružování je určen opět krok a počet prvků, z nichž se vyhodnocuje maximum. *Padding* je obvykle roven hodnotě nula.

Na poslední vrstvě extrakce příznaků dojde k seřazení jednotlivých prvků všech příznakových map do jediné řady (viz seřazení na Obr. 2.3). Vznikne jednorozměrné pole (vektor) seřazených neuronů. Každý neuron je propojen s každým neuronem následující vrstvy. Jedná se o plně propojenou (fully connected) vrstvu. Na úplném

konci CNN je typicky funkce softmax, která normalizuje hodnoty vstupující do této funkce. Výstupem je klasifikace vstupního obrázku v podobě pravděpodobností, do nichž spadá objekt v obrázku. [8]



Obr. 2.6 Znázornění operace sdružování dle maxima přes oblast o velikosti 2×2 , krok je roven 2 a *padding* je roven 0

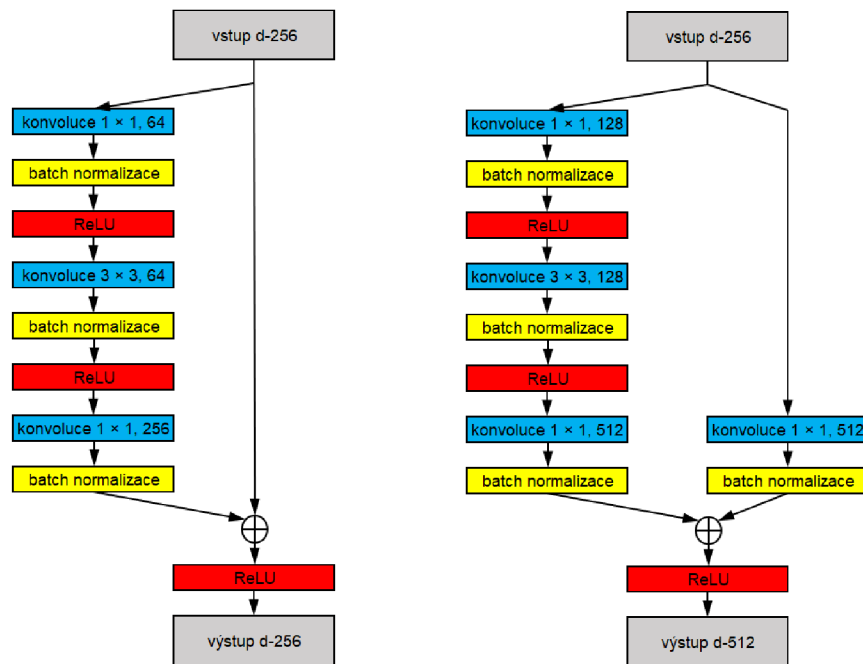
2.1.4 Moderní architektury CNN

Tuto éru moderních architektúr započala CNN AlexNet svým vítězstvím v ročníku 2012 ILSVRC [19]. V dnešní době nejsou architektury CNN jen série konvolučních a sdružovacích vrstev za sebou. Struktury jsou složitější a architektury velmi hluboké nebo široké. Výběr některých známých architektúr je proveden v následujících odstavcích.

Architektura VGG (Visual Geometry Group [30]), mezi nimi nejpopulárnější VGG-16 a VGG-19, je jednoduchá architektura 16 nebo 19 vrstev hluboká, která obsadila druhé místo v ILSVRC 2014. Vrstvy jsou seřazeny sériově za sebou a velikost konvolučního jádra je 3×3 . Myšlenka této architektury je v postupném narůstání počtu konvolučních jader v jedné vrstvě s narůstající hloubkou CNN. Tato architektura nemůže být příliš hluboká, jelikož by hrozil problém mizejícího gradientu (vanishing gradient problem) a s ním související ztráta přesnosti trénování. [1]

Problém mizejícího gradientu u hlubokých CNN řeší **architektura ResNet** (Residual Network [15]) reziduálním blokem. Tato architektura vyhrála první místo v ILSVRC 2015. Zatímco u VGG jsou vrstvy propojené za sebou, přičemž výstup vrstvy předcházející je vstupem vrstvy následující, v reziduálním bloku (Obr. 2.7) je větev složená z konvolucí, *batch* normalizací a ReLU překlenutá druhou větví. Příznakové mapy obou větví jsou posléze sečteny po prvcích, což je v Obr. 2.7 znázorněno symbolem kolečka.

Po konvoluci je provedena *batch* normalizace, která škáluje hodnoty vystupující z konvoluce a tím pomáhá předcházet problému mizejícího gradientu a snižuje potřebný čas na trénování. U této architektury jsou během konvoluce v reziduálním bloku uplatňovány dvě možnosti. Buď reziduální blok zachovává hloubku d vstupní a výstupní vrstvy, toto je případ Obr. 2.7a, anebo zdvojnásobuje hloubku d vrstvy výstupní, jak lze vidět na Obr. 2.7b. Při součtu obou větví musí hloubky obou větví souhlasit, a proto se v pravé větvi v Obr. 2.7b nachází konvoluce 1×1 , která je k tomuto účelu využívána. [1]

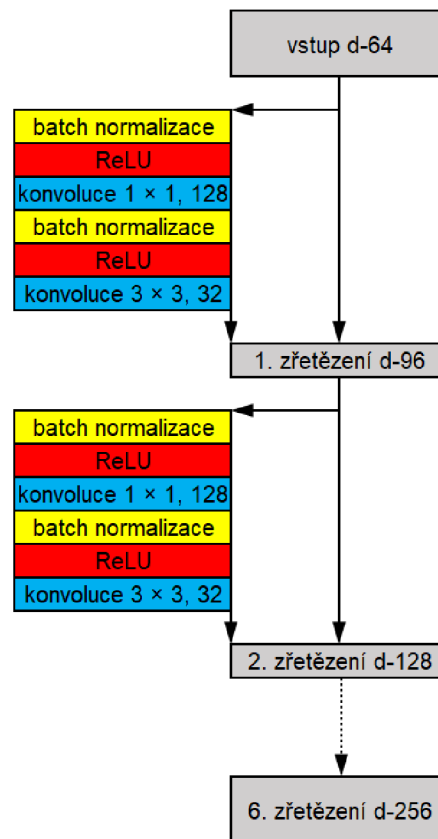


(a) Zachování hloubky vrstvy

(b) Navýšení hloubky vrstvy

Obr. 2.7 Schématické znázornění reziduálního *bottleneck* bloku využívaného v architektuře ResNet-101. Vstupní vrstva má hloubku 256 (d na vstupu a výstupu znamená hloubku vrstvy). V levém případě první konvoluce 1×1 redukuje hloubku vrstvy na 64. Po každé konvoluci následuje *batch* normalizace. Třetí konvoluce zvýší hloubku vrstvy opět na rozměr 256, aby odpovídala vstupní vrstvě a mohlo dojít k součtu obou větví po prvcích, což představuje symbol kolečka. Vpravo je hloubka výstupní vrstvy navýšována, a tak je zde navíc konvoluce 1×1 s *batch* normalizací.

Architektura DenseNet (Dense Convolutional Network [16]) jde v propojování jednotlivých vrstev ještě dále než architektura ResNet. Každá vrstva jednoho *dense* bloku této architektury obdrží příznakové mapy všech vrstev předcházejících a sama sdílí svůj výstup se všemi vrstvami následujícími. Příznakové mapy v tomto případě nejsou sečteny po prvcích, ale jsou zřetězeny (konkatenace). Tempo růstu (*growth rate*) určuje nárůst hloubky vrstvy při zřetězení. Více viz Obr. 2.8. [1]



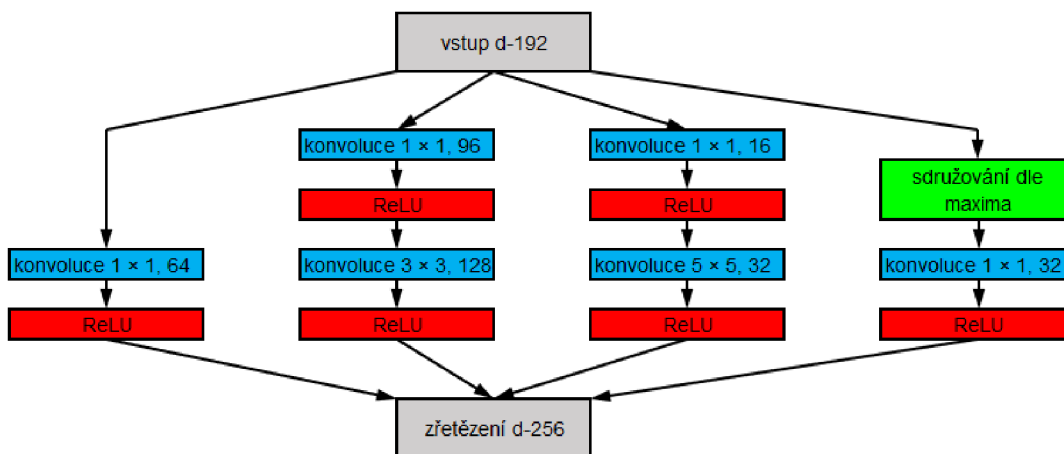
Obr. 2.8 Schéma prvního *dense* bloku architektury DenseNet-201, d je hloubka vrstvy. Tempo růstu (growth rate) je v tomto případě $k = 32$. Celkem má tento *dense* blok 6 větví, které jsou složeny ze dvou *batch* normalizací, dvou ReLU, konvoluce 1×1 a 3×3 . Na obrázku jsou vyznačeny pouze dvě větve z důvodu úspory místa, zbylé čtyři je možno si domyslet namísto tečkované šipky.

Zatímco předcházející dvě architektury se vyznačují svou hloubkou u **architektury Inception Net** [32] je snaha udělat síť širokou. Toho je docíleno paralelním zapojením několika vrstev s rozdílnými konvolučními jádry, jako je tomu na **Obr. 2.9**. Na konci jsou příznakové mapy zřetězeny. Konvoluce 1×1 je zde opět využita pro úpravu hloubky vrstvy, a tedy pro snížení počtu parametrů. [1]

Tato architektura, konkrétně Inception v1, je nazývána jako GoogLeNet a vyhrála ILSVRC 2014 (hned za ní byla architektura VGG, jak již bylo zmíněno). Další vylepšenou variantou je například Inception ResNet [33], která využívá reziduálního bloku – je tedy jakýmsi hybridem architektury ResNet a Inception Net.

Podle webu Papers With Code², kde je možno procházet aktuálně nejlepší výsledky klasifikace obrázků na různých datových sadách, první pozice obsazují v době psaní této práce architektury FixEfficientNet [36], NoisyStudent [39] či BiT-L [18].

² Viz <http://paperswithcode.com>.



Obr. 2.9 Schématické znázornění prvního *inception* modulu architektury Inception v1 (GoogLeNet). Na konci modulu dojde ke zřetězení příznakových map, takže hloubka na výstupu tohoto modulu je $64 + 128 + 32 + 32 = 256$. V pravé větvi je provedeno sdužování dle maxima o velikosti 3×3 a krokem 1.

2.1.5 Transfer learning, matice záměn a přesnost trénování

Využití před-trénované architektury CNN nese název *transfer learning* a nabízí dva přístupy. Prvním a jednodušším je odstřížení první části CNN (extrakce příznaků) od klasifikačních vrstev. Následně je extrakce příznaků využita a výsledné příznaky jsou použity na jakýkoli jiný klasifikátor, např. *support vector machines*. Druhým přístupem je *fine-tuning* (doladění). Spočívá v nahrazení posledních klasifikačních vrstev CNN novými. Posléze je celá CNN trénována na nové datové sadě. U tohoto přístupu je potřeba mít dostatečnou velikost datové sady, jinak by hrozilo přeučení. Přístup *fine-tuning* byl zvolen v této práci, jelikož poskytuje výhodu v tom, že není potřeba budovat novou CNN od začátku, což je proces, který vyžaduje již nabyté zkušenosti. [28], [9]

Matice záměn (confusion matrix) ukazuje predikce CNN pro sady testovacích obrázků. Každá kategorie obsahuje 200 obrázků. Příklad je vidět na [Obr. 2.10](#).

Testovací sada obsahuje štítky se správným názvem kategorie, do níž obrázek této kategorie patří. Toto je v [Obr. 2.10](#) znázorněno pojmem *skutečná kategorie*. Tyto skutečné obrázky jsou předkládány naučené CNN, která je vyhodnocuje. Vyhodnocení, neboli predikce, lze rozdělit do čtyř kategorií. Na diagonále jsou modře počty správně klasifikovaných obrázků – tzv. skutečně pozitivní (true positive, TP). Mimo diagonálu jsou v řádcích hodnoty falešně negativní (false negatives, FN), tedy obrázky, které do dané kategorie patří, ale CNN je nesprávně zařadila do jiné kategorie. Napravo od matice záměn lze vidět TP v prvním sloupci a součty FN ve sloupci druhém.

Falešně pozitivní (false positives, FP) predikce se nachází ve sloupcích mimo diagonálu. V tomto případě CNN nesprávně zařadila obrázek do kategorie, který do ní ve skutečnosti nepatří. Poslední skupinou predikcí jsou skutečně negativní (true negatives, TN). Zde se jedná o obrázky, které nespádaly do dané kategorie a CNN je správně vyhodnotila, že do této kategorie nepatří.

Pod maticí jsou ve dvou řádcích znázorněny opět TP v prvním řádku a součty FP pro jednotlivé kategorie v řádku druhém.

	baby back ribs	beef tartare	churros	donuts	omelette	onion rings	red velvet cake	spaghetti bolognese	sushi	tiramisu		
baby back ribs	188	1	2	1	3	1	0	1	2	1	188	12
beef tartare	3	182	1	3	2	1	1	3	1	3	182	18
churros	4	0	187	3	0	1	1	0	2	2	187	13
donuts	2	0	3	185	1	5	2	0	2	0	185	15
omelette	2	2	2	4	183	0	1	0	5	1	183	17
onion rings	0	1	7	2	1	186	1	2	0	0	186	14
red velvet cake	0	2	2	3	0	0	186	1	0	6	186	14
spaghetti bolognese	2	1	0	1	3	2	2	186	0	3	186	14
sushi	2	4	0	2	2	1	2	0	186	1	186	14
tiramisu	5	1	4	3	2	0	6	0	0	179	179	21
výsledek klasifikace												
	188	182	187	185	183	186	186	186	186	179		
	20	12	21	22	14	11	16	7	12	17		

Obr. 2.10 Matice záměn vzatá z vyhodnocení experimentu pro barevný model RGB

Z matice záměn je vyhodnocena testovací přesnost³. Jedná se o poměr součtu všech skutečně pozitivních predikcí TP (suma diagonály v matici záměn) a celkový počet obrázků testovací sady, který je roven 2000.

2.2 Barevné modely

V počítačové technice je digitální obraz matice čísel. Každý prvek této matice představuje pixel. Obrázek ve stupních šedi je jednovrstvá matice čísel. Jedná se o jednosložkový či jednokanálový obraz (slovem složka a kanál je myšleno totéž a v práci

³ Označuje se také jako přesnost klasifikace nebo celková přesnost.

jsou zaměřovány). Každý pixel nese informaci v podobě čísla. V případě datového typu `uint8` (osmibitová hloubka) používaného v aplikaci MATLAB se jedná o hodnoty v rozmezí 0 až 255. 0 představuje černou, 255 bílou. Barevné obrázky jsou typicky interpretovány třemi kanály – trojrozměrným polem. Vzájemným složením třech kanálů konkrétního pixelu vznikne výsledná barva. Barevný model je způsob interpretace výsledné barvy. Například nejznámější barevný model RGB je složen z červeného, zeleného a modrého kanálu.

Achromatické světlo je bílé, bezbarvé. Jeho vlastností je intenzita. Popisujeme ji ve stupních šedi, které se pohybují od černé až k bílé. Chromatické světlo se na elektromagnetickém spektru pohybuje v rozmezí přibližně od 400 nm (fialová) do 700 nm (červená). Veličiny kvalitativně popisující zdroj chromatického světla jsou zář (radiance), jas (luminance) a světlost (brightness⁴). Zář je celkové množství energie, které vyzařuje světelný zdroj. Jas vyjadřuje množství energie, které pozorovatel přijme ze světelného zdroje. Světlost je vizuální vjem, který je subjektivní a je nemožno jej měřit. [11]

V lidském oku se nachází čípky, které umožňují barevné vidění. Přibližně 65 % čípků jsou vnímavé na červenou část spektra, 33 % na zelenou a pouze 2 % na modrou. Nicméně modré čípky jsou nejvíce citlivé na světlo. Z tohoto biologického konceptu tří typů receptorů vychází základní barvy barevného modelu RGB: červená, zelená a modrá. [11]

Veličiny sloužící k odlišení jednotlivých barev jsou: světlost (brightness), odstín (hue) a sytost (saturation). O světlosti již bylo pojednáno. Odstín představuje převažující vlnovou délku světla, kterou vnímá pozorovatel. Sytost odkazuje k čistotě barvy. Jedná se o množství bílé barvy smíchané s daným odstínem. Čisté barvy mají 100% sytost. Odstín a sytost dohromady představují barevnost (chromaticity). Tři složky základních barev, jež jsou potřebné pro vytvoření jakékoli barvy, se nazývají trichromatické složky. [11]

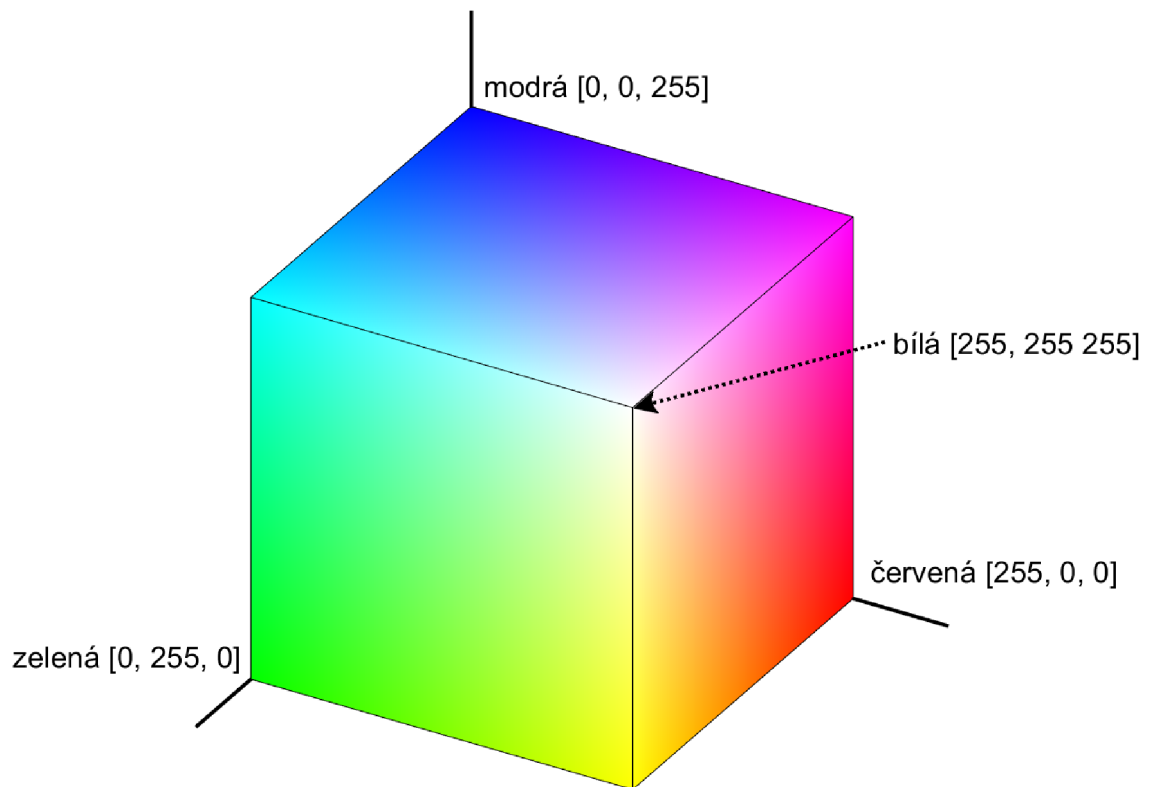
2.2.1 RGB barevný model

Barevný model je znázorněn krychlí v kartézském souřadnicovém systému, kde černá barva je v počátku souřadnicového systému, bílá barva přes uhlopříčku krychle. Základní barvy jsou umístěny na jednotlivých osách. Jejich doplňkové barvy (azurová, purpurová a žlutá) pak ve zbylých vrcholech. Takto je to vidět na Obr. 2.11.

Jedná se o model založený na aditivním míchání barev. To znamená, že do černé barvy je možno postupným přidáváním základních barev namíchat jednotlivé barvy.

⁴ *Brightness* je spojován ve zobrazovací technologii také s pojmem jas (subjektivní jas), jedná se ale o bezrozměrnou veličinu. Alternativním anglickým názvem je pojem *lightness*, přičemž pojem *brightness* je vztahován k svítícím objektům a *lightness* k nesvítícím objektům. [31]

V případě obrázku, jehož každý pixel je popsán 24 bity (true color), je každá složka RGB modelu v obrázku zastoupena osmi bity. Tato vlastnost se nazývá barevná hloubka. Celkový počet barev v takovémto obrázku je $2^{24} = 16\,777\,216$. Pokud je v MATLABu obrázek zastoupen datovým typem `uint8`, každá složka pixelu nabývá hodnoty 0–255. Tento barevný model je vhodný pro obrazovky a kamery. Nicméně jeho nevýhodou je závislost na zařízeních. Pokud zachytíme stejný obraz na dvou různých kamerách, hodnoty RGB se budou lišit. [35], [37]



Obr. 2.11 Znárodnění barevného modelu RGB. V počátku souřadnicového systému $[0, 0, 0]$ je černá barva. Bílá barva se nachází na souřadnici $[255, 255, 255]$. Spojnice (úhlopříčka) těchto dvou bodů představuje stupně šedi. Na jednotlivých osách R , G , B jsou základní barvy, doplňkové barvy se nachází na zbylých třech vrcholech a vzniknou vždy smícháním dvou základních barev: azurová z modré a zelené, purpurová z červené a modré a žlutá z červené a zelené. Vytvořeno pomocí aplikace MATLAB.

2.2.2 HSV barevný model

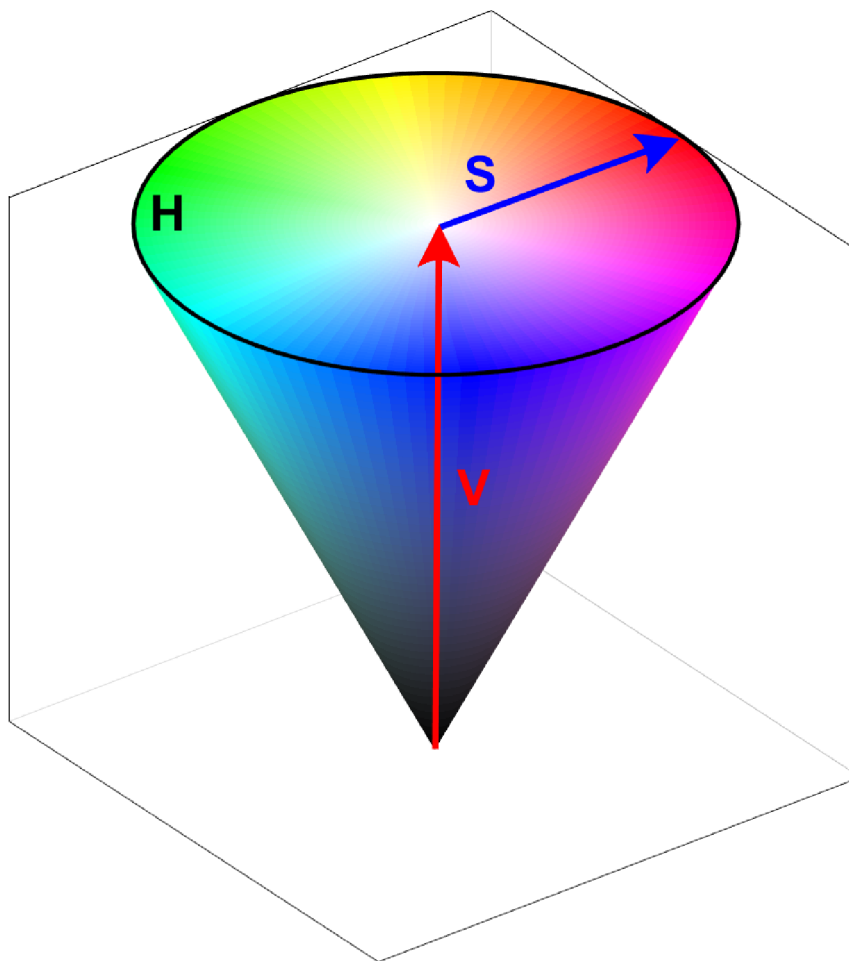
Model HSV je oproti RGB modelu intuitivní pro interpretaci barev člověkem. Je tedy hojně využíván pro míchání barev v grafických editorech. Tento model je složen ze složek odstínu H , sytosti S a relativní světlosti barvy (value, V). Relativní světlost udává, jak moc se odstín odlišuje od černé [31], neboli kolik světla daný odstín emituje. Geometricky je tento model znázorněn šestibokým jehlanem stojícím na svém hlavním vrcholu, tedy základnou vzhůru. Hlavní vrchol představuje černou barvu,

výška jehlanu reprezentuje relativní světlost, jež počíná ve vrcholu (černá) a končí v základně (bílá). Sytost je kolmá vzdálenost od výšky jehlanu (relativní světlosti). Čisté základní a doplňkové barvy se nachází ve vrcholech základny. Tento model je možno reprezentovat také geometrickými transformacemi šestibokého jehlanu – kuželem či válcem. Převod z RGB modelu je dán nelineární transformací pomocí algoritmu (v MATLABu příkaz `rgb2hsv`):

$$\begin{aligned}
 V &= \max(R, G, B) \\
 X &= \min(R, G, B) \\
 S &= \frac{V - X}{V} \\
 r &= \frac{V - R}{V - X} \\
 g &= \frac{V - G}{V - X} \\
 b &= \frac{V - B}{V - X} \\
 H &= \begin{cases} \text{nedefinován} & \text{pro } V = X \\ 5 + b & \text{pro } R = V \text{ a } G = X \\ 1 - g & \text{pro } R = V \text{ a } G \neq X \\ 1 + r & \text{pro } G = V \text{ a } B = X \\ 3 - b & \text{pro } G = V \text{ a } B \neq X \\ 3 + g & \text{pro } R = X \\ 5 - r & \text{v ostatních případech} \end{cases}
 \end{aligned} \tag{2.1}$$

Jedná se o normalizované hodnoty HSV modelu pohybující se v intervalu $\langle 0, 1 \rangle$. V případě geometrické interpretace (jehlan, kužel, hranol) je odstín znázorněn úhlem průvodiče v základně. 0° pro červenou barvu, 120° pro zelenou, 240° pro modrou. Doplňkové barvy pak leží na 60° (žlutá), 180° (azurová) a 300° (purpurová). Tyto úhly jsou dány konvencí. HSV model znázorněný kuželem je zobrazen na [Obr. 2.12](#).

Zde je možno vidět nevýhodu v diskontinuitě modelu, kdy 359° a 0° představuje téměř totožný odstín červené. Toto téma je dále rozebíráno v praktické části v kapitole [3.5](#). Speciální pozornost je potřeba také věnovat singulárnímu případu, kdy $R = G = B$. Na ose relativní světlosti není odstín definován, v algoritmu nastává dělení nulou. MATLAB odstín H v tomto případě stanovuje jako nulu. [\[31\]](#), [\[37\]](#)



Obr. 2.12 Znárodnění barevného modelu HSV. Odstín H je dán úhlem průvodiče v základně kuželu, sytost S a relativní světlost V je v rozmezí $(0, 1)$.

2.2.3 XYZ barevný prostor

V roce 1931 definovala Mezinárodní komise pro osvětlování CIE (Commission internationale de l'éclairage) z velkého množství experimentálně naměřených dat tzv. standardního pozorovatele. Popisuje průměrnou citlivost lidského oka na barvy. Standardizované vlnové délky pro tři primární barvy jsou: modrá = 435,8 nm, zelená = 546,1 nm a červená = 700 nm. Prostor⁵ CIE 1931 XYZ (dále jen XYZ) byl definován taktéž v roce 1931. Kanál Y představuje jas barvy, Z přibližně odpovídá množství modré v obrázku, ale není identický se složkou B v RGB modelu. X není analogický s žádnou barvou, ale leží podél kolmé osy na složky Y a Z . Pro převod z barevného modelu RGB do modelu XYZ⁶ je nutno nejprve převést barevný

⁵ V kontextu barevných modelů vytvořených CIE se pojednává spíše o termínu barevný prostor než barevný model, ačkoli podle [11] se jedná o pojmy totožné. Článek [17] pojednává i o termínu barvový prostor.

⁶ Tento postup platí pro standardní osvětlení D65.

model do standardního sRGB modelu [21]:

$$\begin{aligned}
 R_{\text{sRGB}} &= \begin{cases} R/12,92 & \text{pro } R \leq 0,04045 \\ [(R + 0,055)/1,055]^{2,4} & \text{v ostatních případech} \end{cases} \\
 G_{\text{sRGB}} &= \begin{cases} G/12,92, & \text{pro } G \leq 0,04045 \\ [(G + 0,055)/1,055]^{2,4} & \text{v ostatních případech} \end{cases} \\
 B_{\text{sRGB}} &= \begin{cases} B/12,92, & \text{pro } B \leq 0,04045 \\ [(B + 0,055)/1,055]^{2,4} & \text{v ostatních případech} \end{cases}
 \end{aligned} \tag{2.2}$$

Hodnoty R , G , B nutno brát v rozsahu $\langle 0, 1 \rangle$. Souřadnice X , Y a Z jsou poté spočítány pomocí:

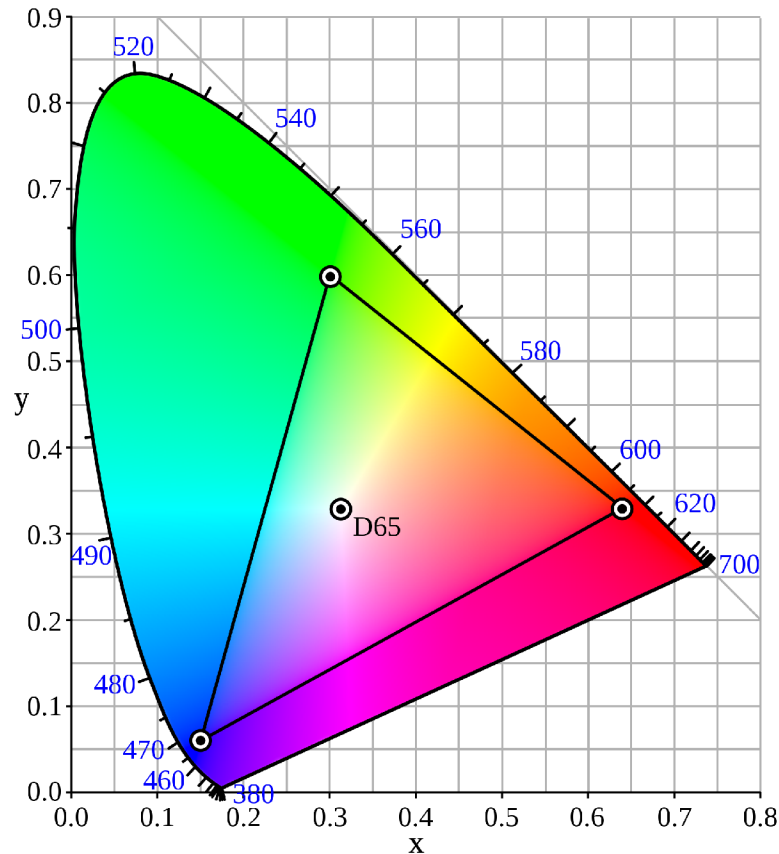
$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,4124564 & 0,3575761 & 0,1804375 \\ 0,2126729 & 0,7151522 & 0,0721750 \\ 0,0193339 & 0,1191920 & 0,9503041 \end{bmatrix} \begin{bmatrix} R_{\text{sRGB}} \\ G_{\text{sRGB}} \\ B_{\text{sRGB}} \end{bmatrix} \tag{2.3}$$

V MATLABu je transformace z barevného modelu RGB provedena pomocí příkazu `rgb2xyz`. Z barevného prostoru XYZ vychází barevný prostor xyY, který užívá chromatických souřadnic x a y vypočtených ze souřadnic X , Y , Z :

$$\begin{aligned}
 x &= \frac{X}{X + Y + Z} \\
 y &= \frac{Y}{X + Y + Z} \\
 z &= \frac{Z}{X + Y + Z}
 \end{aligned} \tag{2.4}$$

Pomocí dvou těchto souřadnic možno dopočítat třetí z rovnice: $x + y + z = 1$. Složka jasu barvy Y prostoru xyY je totožná se složkou Y prostoru XYZ. Hodnoty x , y je možno zobrazit v chromatickém diagramu, což je užitečný nástroj, který významným způsobem zpřehledňuje práci s barvami. Poskytuje názorný náhled všech viditelných barev a je možné jej vidět na [Obr. 2.13](#).

S těmito prostory se používá pojem gamut, což je množina barev zobrazitelná v určitém barevném prostoru nebo na daném zařízení. Na [Obr. 2.13](#) je znázorněn gamut sRGB v podobě trojúhelníku. [11], [17], [37]



Obr. 2.13 Chromatický diagram CIE xy s gamutem sRGB a standardním osvětlením D65 (bílý bod). Osy představují chromatické souřadnice x a y , hraniční křivka tvaru podkovy se nazývá spektrální locus a ukazuje vlnové délky čistých spektrálních barev. Čára spojující koncové body nese jméno purpurová linie a jsou na ní zobrazeny nespektrální barvy. Převzato z [25].

2.2.4 LAB barevný prostor

Barevný prostor CIE 1976 $L^*a^*b^*$ (dále jen LAB) je prostor perceptuálně uniformní. To znamená, že číselná vzdálenost mezi dvěma rozdílnými odstíny barvy v tomto barevném prostoru je vnímána pozorovatelem subjektivně jednotně. Složka L^* se pohybuje v rozmezí $\langle 0, 100 \rangle$ a představuje světlost. Černá barva je zastoupena hodnotou 0 a bílá hodnotou 100. Množství červených a modrých barevných tónů je specifikováno kanálem a^* a množství žlutých nebo modrých tónů je dáno kanálem b^* . Kanály a^* a b^* se pohybují v intervalu buď $\langle -100, 100 \rangle$, anebo $\langle -128, 127 \rangle$.

Transformaci lze provést v MATLABu příkazem `rgb2lab` a lze ji docílit také z barevného prostoru XYZ použitím následujících rovnic:

$$\begin{aligned}x_{\text{ref}} &= \frac{X}{X_{\text{ref}}} \\y_{\text{ref}} &= \frac{Y}{Y_{\text{ref}}} \\z_{\text{ref}} &= \frac{Z}{Z_{\text{ref}}}\end{aligned}\tag{2.5}$$

$X_{\text{ref}} = 0,95047$, $Y_{\text{ref}} = 1$ a $Z_{\text{ref}} = 1,08883$ jsou hodnoty pro standardní osvětlení D65 získané z MATLABu příkazem `whitepoint('d65')`. Pomocí chromatických souřadnic x a y , jejichž výpočet (2.4) je zmíněn v předchozí podkapitole 2.2.3, lze určit polohu bílého bodu D65 v chromatickém diagramu (Obr. 2.13): $x = 0,31$ a $y = 0,33$. Dále se při výpočtu L^* , a^* , b^* postupuje následovně:

$$\begin{aligned}f_x &= \begin{cases} \sqrt[3]{x_{\text{ref}}} & \text{pro } x_{\text{ref}} > 0,008856 \\ 7,787 \cdot x_{\text{ref}} + 16/116 & \text{v ostatních případech} \end{cases} \\f_y &= \begin{cases} \sqrt[3]{y_{\text{ref}}} & \text{pro } y_{\text{ref}} > 0,008856 \\ 7,787 \cdot y_{\text{ref}} + 16/116 & \text{v ostatních případech} \end{cases} \\f_z &= \begin{cases} \sqrt[3]{z_{\text{ref}}} & \text{pro } z_{\text{ref}} > 0,008856 \\ 7,787 \cdot z_{\text{ref}} + 16/116 & \text{v ostatních případech} \end{cases}\end{aligned}\tag{2.6}$$

Nyní už stačí dosadit do výpočtu jednotlivých kanálů:

$$\begin{aligned}L^* &= 116 \cdot f_y - 16 \\a^* &= 500 \cdot (f_x - f_y) \\b^* &= 200 \cdot (f_y - f_z)\end{aligned}\tag{2.7}$$

Podání barvy na vytisknutém papíru nebo zobrazení na dvou různých monitorech se liší. Tato vlastnost je nazývána jako závislost na zařízení. Barevný prostor XYZ společně s prostorem LAB představují prostory, které jsou nezávislé na zařízení, a jsou proto využívány v systému správy barev pro převod barevných modelů mezi zařízeními. Barvy prostoru LAB nelze přímo zobrazit, je potřeba transformace do jiného prostoru. Gamut prostoru LAB zahrnuje celé barevné spektrum. [37], [11]

2.2.5 YIQ barevný model

Model YIQ⁷ je definován standardem NTSC pro analogový televizní signál. Je úzce spjat s rozvojem barevné televize. Složka Y ukládá jasovou informaci⁸. Lépe řečeno jde o světlost (subjektivní jas), jelikož tyto modely jsou používány ve zobrazovací technologii. Kanál I znázorňuje přibližné množství modrých nebo oranžových barevných tónů, zatímco Q představuje přibližné množství zelených nebo purpurových tónů. I znamená *in-phase* a Q *quadrature*, což jsou pojmy související s modulací signálu. Tento model separuje jasovou složku od barevnosti, takže stejný signál může být použit jak pro černobílé televize, tak pro barevné. [22]

V MATLABu je transformace z RGB modelu provedena příkazem `rgb2ntsc`, přičemž je využito vzorce:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 \\ 0,596 & -0,274 & -0,322 \\ 0,211 & -0,523 & 0,312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.8)$$

Za R , G , B je nutno dosadit hodnoty nikoli v rozsahu $\langle 0, 255 \rangle$, ale v normalizovaném rozsahu $\langle 0, 1 \rangle$. [37]

2.2.6 YCbCr barevný model

Barevný model YCbCr je široce využíván u digitálního videa a fotografie. Složka Y ukládá informaci jasu barvy. Barevná informace je uchována v kanálech Cb a Cr , jež odlišují modrou a červenou složku od referenční hodnoty. V MATLABu je transformace provedena příkazem `rgb2ycbcr` nebo pomocí následujícího vzorce, který je možno nalézt v MATLABu otevřením funkce `rgb2ycbcr`:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 65,481 & 128,553 & 24,966 \\ -37,797 & 74,203 & 112,000 \\ 112,000 & 93,786 & -18,214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (2.9)$$

kde Y je v rozsahu $\langle 16, 235 \rangle$ a Cb a Cr v intervalu $\langle 16, 240 \rangle$ za předpokladu, že se jedná o datový typ pole `uint8`. Hodnoty R , G , B nutno opět zadávat v rozmezí $\langle 0, 1 \rangle$. [37]

Všechny výpočty uvedené v této podkapitole o barevných modelech je možno ověřit v souboru: `\prilohy\transformace.xlsx`. Barevné modely lze rozdělit do dvou kategorií podle použití:

⁷ V příloze práce je pojmenován jako NTSC.

⁸ U modelu YIQ a YCbCr je v souvislosti s televizní technikou používáno pro jas anglické slovíčko *luma* namísto *luminance* [27].

- barevné modely závislé na zařízení, patří mezi ně RGB, YIQ, YCbCr a HSV,
- barevné prostory nezávislé na zařízení, mezi něž se řadí XYZ a LAB.

2.3 Poznatky o vlivu barevných modelů na chování CNN

Nejvíce článků zkoumajících vliv barevných modelů na chování CNN je z roku 2019. Z toho lze usuzovat, že toto téma je v současné době živé. Způsobů, jak zkoumat chování CNN, je totiž mnoho. Je to nejen samotná volba barevného modelu, která má vliv na přesnost klasifikace, ale také ladění hyperparametrů, volba trénovacího algoritmu (ADAM, SGD atd.), volba architektury CNN, volba datové sady a další.

V článku [2] autoři porovnávají barevné modely HSV a CMYK pro úlohu detekce make-upu. Využívají těchto datových sad: YouTube Makeup, Virtual Makeup Database, Makeup In the Wild, Facial Cosmetics Database a Makeup Induced Face Spoofing. Binární klasifikaci realizují pomocí jednoduché CNN. Autoři zmiňují důležitost modelu HSV pro jeho schopnost separovat složku odstínu, nasycení a relativní světlosti, což se ukazuje jako výhoda v porovnání s modelem RGB v úloze tohoto typu. Transformace do modelu CMYK je provedena pomocí profilu ICC Uncoated-FOGRA29. Závěr je takový, že model CMYK si vedl obecně lépe oproti modelu HSV. Ten dokázal obstát pouze v detekci absence make-upu.

Zdroj [5] zkoumá čtyři datové sady: PersonFinder, CIFAR-10, CIFAR-100 a FlickrScene. U první zmíněné datové sady se vliv barevných modelů ukázal jako zanedbatelný. V přesnosti klasifikace vyhrála jasová složka, takže autoři tohoto článku uvádějí, že daný model je nezávislý na barvě. CNN zkrátka nadbytečnou informaci v podobě barvy při tréninku na těchto sadách nepotřebuje. K podobnému závěru došli autoři článku [23], kteří použili datovou sadu MIT Pedestrian. Důvod je intuitivní. Obě datové sady obsahují fotky chodců. Každý chodec má jinou barvu oblečení, je zde tedy velká variabilita barev a CNN se při tréninku „soustředí“ spíše na tvary než na barvy. Naopak největší roli sehrála v článku [5] barevná informace u datové sady FlickrScene, která obsahuje různé druhy scénérií. Zde je například důležité rozlišit zasněžené kopce od písčitých. I pro člověka jsou tyto aspekty těžko rozlišitelné, pokud chybí barva. U městské scénérie ale opět vliv barevné informace nebyl tak dramatický. U datové sady CIFAR-100 se tyto závěry potvrdily: barva se ukázala jako druhořadá u kategorií aut, stromů a květin. Autoři také komentují horší přesnost testování modelu HSV. Přisuzují ji ztrátě některých tvarů vlivem diskontinuity transformace z modelu RGB do HSV. A tvrdí, že transformace do modelu HSV zhoršuje kvalitu tvarů. O této vlastnosti se pojednává v podkapitole 3.5 a na obrázku Obr. 3.6.

V úloze detekce včel z videa spatřují autoři článku [7] výhodu v trénování jednoho kanálu (stupně šedi vytvořené z RGB nebo kanál H modelu HSV) v menší výpočetní náročnosti. Lze je použít v neobvyklých světelných podmínkách, při umělém osvětlení či při atypickém vyvážení bílé digitální kamery. Datovou sadu, kterou autoři vytvořili, volně poskytují k užití. Dále podtrhují důležitost barev při trénování CNN v úloze detekce včel.

Práce [10] sledovala vliv barevných modelů RGB, LAB, LUV, dRdGdB, Y0102 a HSV na algoritmy detekce hran. Tyto algoritmy jsou mezikrokem v aplikacích počítačového vidění pro detekci nebo rozpoznání objektů. Autoři použili datovou sadu BSDS500 (500 obrázků přírody) a PASCAL Context (kategorie jako vidlička, obloha, cesta, klávesnice, voda atp.) a zjistili, že bez ohledu na použitou sadu, vyhrály perceptuálně uniformní modely (LAB a LUV).

Tvůrci článku [13] dokazují, že různé kategorie obrázků datové sady CIFAR-10, CIFAR-100, SVHN a ImageNet jsou lépe reprezentovány v různých barevných modelech. Kombinují také jednotlivé barevné modely na vstupu CNN. To znamená, že namísto třech kanálů poskytují CNN šest kanálů (v případě kombinace dvou barevných modelů), devět kanálů (kombinace tří barevných modelů) atd. Tento způsob se nazývá multi-barevný model (multi-color space) a dosahuje vyšší přesnosti než u barevných modelů trénovaných individuálně. Autoři zmiňují důležitost vlivu hloubky CNN pro přesnost klasifikace. Upozorňují však na problém mizejícího gradientu a s ním související ztrátu přesnosti trénování. Zabývají se myšlenkou, že nikoli větší hloubka CNN, ale větší šířka může zlepšit přesnost. Tuto hypotézu dokazují experimentem, kdy použili sedm barevných modelů: RGB, YIQ, LAB, HSV, YUV, YCbCr a HED. Každý tento model trénovali simultánně na jedné architektuře DenseNet a ve vrstvě Dense Layer pak tyto výstupy z jednotlivých sítí zkombinovali, čímž zabránili přetrénování a zároveň při mnohem menším počtu parametrů dokázali v přesnosti konkurovat současným trendům. Autoři zkoušeli také přesnosti jednotlivých barevných modelů na datové sadě CIFAR-10. Nejlépe dopadl model LAB, ale rozdíly v přesnosti klasifikace v porovnání s ostatními barevnými modely byly nepatrné.

Multi-barevnými modely zmíněnými v předchozím odstavci se zabývají také autoři článku [35]. Spojují dohromady modely RGB, HSV a LAB zřetězením kanálů. Na vstup CNN tedy poskytují devítikanálový obraz. V tomto článku je zmíněn důležitý pozitivní vliv nelinearity transformace na přesnost klasifikace. Multi-barevné modely zvyšují nelinearitu, a tedy snižují nedostatky RGB modelu trénovaného samostatně. Autorům se podařilo zvýšit přesnost klasifikace touto metodou.

V práci z roku 2020 [9] autoři zmiňují, že se prozatím neseťkali se studiem vlivu barevných modelů na metodu *transfer learning*. Také zmiňují časovou náročnost transformace do jednotlivých barevných modelů během trénování, a doporučují proto pro-

vést transformaci ještě před zahájením samotného trénování. Jejich metodika spočívala ve využití třech architektur CNN: Inception-V3, ResNet a MobileNet. Tyto architektury použili jako extraktory příznaků – využili tedy první část CNN. Posléze extrahované příznaky použili pro přeučení klasifikačních vrstev pěti různých podsad datové sady Kaggle (kočky vs. psi). Každá tato podsada byla v jiném barevném modelu: RGB, LAB, posléze trojice jednotlivých kanálů z barevného prostoru LAB, tedy LLL, AAA, BBB. Závěr je takový, že nejen barevný model, ale i architektura CNN ovlivňuje přesnost. Například MobileNet neřeší rozdíly v barevné informaci, ale zabývá se pouze prostorovou informací obrazu, tedy tvary. Podle článku [5] lze říci, že tato architektura je nezávislá na barevné informaci a postačí ji ke klasifikaci jednokanálové modely, čímž se zvýší výkonnost klasifikace (sníží se trénovací čas).

Autoři článku [9] doporučují u architektury ResNet prozkoumat vliv trénování celé architektury ResNet v jiných barevných modelech a na datové sadě, u níž je klasifikační úloha závislá na barevné informaci, což může být podle zdroje [5] sada FlickrScene. U architektury ResNet vyhrál v přesnosti barevný model RGB a u Inception-V3 zvítězil v přesnosti barevný model LAB. Autoři dodávají, že aby mohli generalizovat výsledky své práce, potřebovali by použít rozdílné datové sady.

Kapitola 3

Praktická část

Tato kapitola se věnuje experimentu se šesti barevnými modely RGB, HSV, XYZ, LAB, YIQ a YCbCr a architekturou ResNet-101. Nejprve je experiment blíže přiblížen, je popsána datová sada Food-101, s níž bylo pracováno. Posléze je proveden inicializační experiment, na kterém byla testována volba parametrů pro hlavní experiment, jenž následuje po inicializačním experimentu. Poslední podkapitola se věnuje barevnému modelu HSV.

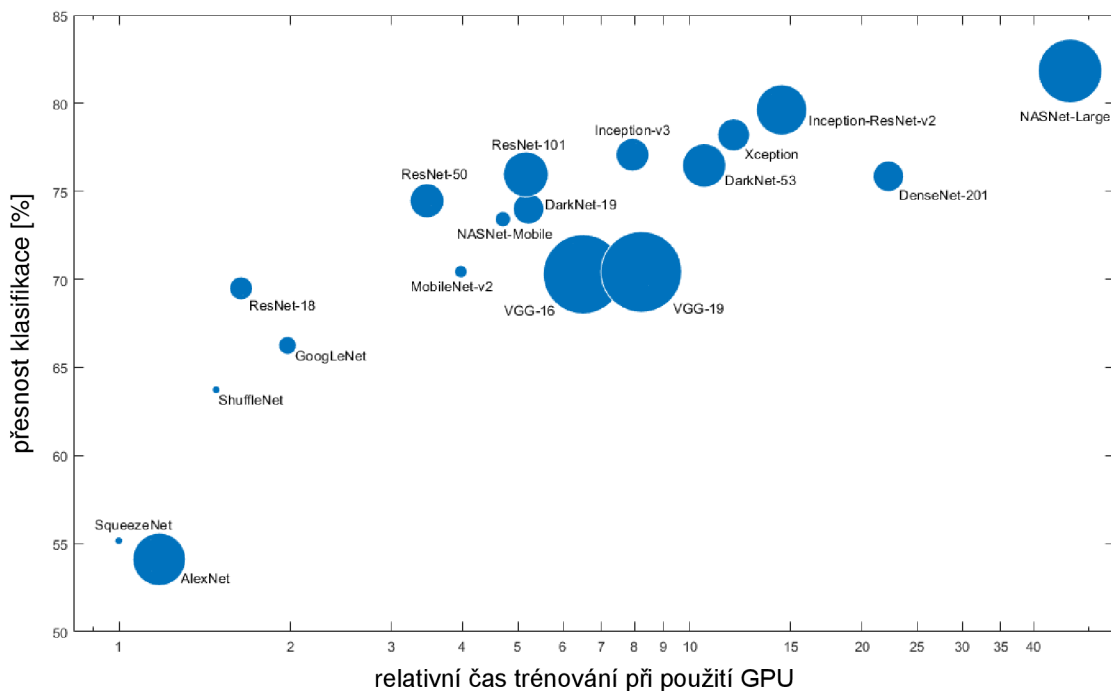
3.1 Popis experimentu

Experiment byl prováděn v aplikaci MATLAB R2019a. Skripty je možno nalézt v příloze ve složce: `\prilohy\skripty`. Skript `main.m` obsahuje hlavní skript pro spuštění experimentu. Funkce `transformFcn.m` provádí transformaci do jednotlivých barevných modelů a `saveTrainingInfo.m` ukládá čistý trénovací čas experimentu do proměnné `data.mat`. Tyto proměnné se nachází ve složkách každého barevného modelu a krom čistého trénovacího času je zde uložena také přesnost každého opakování. Experiment byl realizován na školním počítači Beast s grafickou kartou NVIDIA GeForce GTX TITAN X s 12 GB pamětí, osmijádrovým procesorem AMD FX-8370 (pracovní frekvence jader 4 GHz) a pamětí RAM 32 GB.

Pro experiment byla zvolena architektura ResNet-101. Název napovídá, že je 101 vrstev hluboká. Byla zvolena jako kompromis mezi potřebným trénovacím časem a přesností, jak je to vidět na [Obr. 3.1](#). Tato architektura využívá tzv. *bottleneck designu*, o němž bylo pojednáno v kontextu [Obr. 2.7](#). [15]

Zvolená metoda před-trénované architektury CNN se jmenuje *transfer learning* s přístupem *fine-tuning*, který je vhodný v tom, že není potřeba budovat novou architekturu CNN od začátku, což je proces, který vyžaduje již získané zkušenosti. Podle [26] postačí nahradit v architektuře ResNet-101 vrstvu třetí od konce. Jedná se o vrstvu

s názvem `fc1000` (1000 kategorií, na které byla původně CNN trénovaná) vrstvou se jménem `fc10` (pro 10 nových kategorií, na něž je potřeba CNN trénovat). Označení vrstvy je libovolné, jediný parametr, který je potřeba upravit v příkazu `fullyConnectedLayer`, je první parametr `outputSize` na hodnotu podle počtu kategorií. Další vrstvou, jež je potřeba nahradit, je vrstva poslední a nese označení `ClassificationLayer_predictions`.



Obr. 3.1 Architektury CNN nabízené MATLABem, na vodorovné ose je čas potřebný k tréninku dané architektury vztažený k nejrychlejší architektuře. Přesný čas závisí na použitém hardwaru a velikosti *mini-batch*. Svislou osu představuje přesnost klasifikace v procentech a velikosti bublin znázorňují velikost architektury v MB. Převzato z [28].

3.2 Datová sada

Kritériem pro volbu datové sady byl dostatečný počet obrázků v přiměřeném rozlišení. ResNet-101 potřebuje na svém vstupu obrázky o velikosti 224×224 . Datová sada CIFAR-100 byla zavrhnuta, jelikož obsahuje obrázky o velikosti 32×32 . Po delším pátrání byla zvolena datová sada Food-101¹ [3]. Jedná se o datovou sadu, která obsahuje obrázky 101 kategorií jídel, jež byly získány z aplikace Foodspotting (tato aplikace ale bohužel v květnu roku 2020 zanikla). Každá kategorie Food-101 obsahuje 1000 obrázků.

¹ Dostupné na <https://www.kaggle.com/dansbecker/food-101>.

Pro experimenty bylo zvoleno 10 kategorií tak, aby tyto kategorie nebyly obsaženy v kategoriích před-trénované CNN v MATLABu. Zároveň podle [26] je doporučeno zvolit kategorie obrázků s podobnými rysy, jako je tomu u kategorií před-trénované CNN. To znamená zvolit obrázky podobného typu. Všechny zabudované CNN v MATLABu² jsou naučeny na 1 000 kategoriích datové sady ImageNet. Příklady kategorií jsou zvířata, nábytek, předměty, dopravní prostředky, ovoce atp. Z kategorií Food-101 byly zvoleny jednodušší ve smyslu rozpoznatelnosti. Konkrétně se jedná o: baby back ribs (žebírka), beef tartare (hovězí tatarák), churros, donuty, omeleta, cibulové kroužky (onion rings), dort red velvet, špagety bolognese, sushi a tiramisu.

Datová sada byla pročištěna, jelikož některé obsahovaly tzv. parazitní obrázky. Tyto obrázky očividně nespádaly do dané kategorie. Následně byly smazány také duplicitní obrázky pomocí aplikace Awesome Duplicate Photo Finder³. Smazané obrázky byly doplněny obrázky z podobné datové sady UPMC Food-101⁴ [38]. Tato datová sada obsahuje stejné kategorie jako Food-101, ale obrázky jsou získány z Google Image.

Datová sada byla rozdělena na 800 trénovacích a 200 testovacích obrázků. Pro validaci byla použita stejná sada jako pro testování. Obrázky datové sady jsou zmenšeny na potřebnou velikost 224×224 pomocí příkazu `augmentedImageDatastore` a transformace do jednotlivých modelů je provedena příkazem `transform` přes funkci `transformFcn.m`. V experimentech dochází v každé epoše k náhodnému zamíchání trénovacích dat.

3.3 Inicializační experiment

Prvotním experimentem bylo třeba nalézt vhodné nastavení hyperparametrů a potřebný počet epoch pro trénování. Bylo zvoleno 100 epoch (tj. 20 000 iterací) pro každých šest barevných modelů (RGB, HSV, XYZ, LAB, YIQ a YCbCr). Míra učení `InitialLearnRate` byla nastavena na velikost 0,01 a `MiniBatchSize` byl nastaven na 40. Všechny hyperparametry, které byly nastaveny v `trainingOptions` je možno vidět v Tab. 3.1.

Tímto experimentem byly ověřeny dané hyperparametry, pro následující hlavní experiment postačí počet epoch nastavit na 30. Výsledky je možno najít ve složce: `\prilohy\1_inicializacni_experiment`⁵. Výstupem tohoto experimentu je tré-

² Je možno tyto architektury stáhnout a nainstalovat přes doplňky Add-On, je zároveň potřeba mít Deep Learning Toolbox.

³ Dostupné na <https://www.duplicate-finder.com>.

⁴ Dostupné na <http://visiir.lip6.fr>.

⁵ Model YIQ je v přílohách označován jako NTSC.

novací křivka `trainingProgress.png` a matice záměn `cm.png`. Dále testovací přesnost a trénovací čas, které jsou uloženy v proměnné `data.mat`. Shrnutí výsledků viz [Tab. 3.2](#).

Tab. 3.1 Nastavení hyperparametrů pro inicializační experiment

<code>solverName</code>	<code>sgdm</code>
<code>InitialLearnRate</code>	<code>0.01</code>
<code>MaxEpochs</code>	<code>100</code>
<code>MiniBatchSize</code>	<code>40</code>
<code>Shuffle</code>	<code>every-epoch</code>
<code>ValidationFrequency</code>	<code>200</code>

Tab. 3.2 Shrnutí výsledků inicializačního experimentu

	RGB	HSV	XYZ	LAB	YIQ	YCbCr
přesnost testování	91,70 %	84,90 %	91,00 %	90,25 %	90,55 %	91,25 %
doba trénování	16.16 h	16.14 h	17.54 h	20.26 h	17.11 h	16.32 h

Z výše uvedené tabulky je patrné, že RGB model si vedl nejlépe, následují modely YCbCr, XYZ, YIQ, LAB a HSV. HSV model ale potřeboval nejkratší dobu na trénování a zvítězil nad modelem RGB, navzdory tomu, že zde musela být provedena transformace během tréninku. Přesnosti barevných modelů XYZ, LAB, YIQ a YCbCr se pohybují v rámci jednoho procentního bodu. Je třeba ale věnovat pozornost přesnosti barevného modelu HSV. Ta je nízká v porovnání s ostatními barevnými modely. Bylo tedy nutné otestovat i jiné nastavení míry učení pro tento barevný model.

Byly provedeny tři experimenty s různou mírou učení: $5e-3$ (tj. 0,005), $5e-4$ (0,0005) a $5e-5$. Zvoleno bylo 50 epoch (10 000 iterací). Poté byly otestovány tyto míry učení: $1e-1$, $1e-3$, $1e-4$ a $1e-5$ na 100 epochách (20 000 iterací). Nakonec se ukázalo, že pro poslední zmíněnou míru učení nemusí být 100 epoch dostatečných, a tak byl proveden ještě jeden experiment na 300 epochách (60 000 iterací). Výsledné přesnosti jsou shrnuty v [Tab. 3.3](#). V příloze je možno najít výsledky ve složce: `\prilohy\2_ladeni_hsv`.

Pro míru učení $1e-5$ se počet epoch ukázal jako nedostatečný, nicméně i při trojnásobném počtu epoch testovací přesnost nepřekonala výsledek pro nastavení míry učení na $1e-2$. Podle [Tab. 3.3](#) tedy není problém v nastavení míry učení.

Tab. 3.3 Přesnosti testování pro různé nastavení míry učení

míra učení	počet epoch	testovací přesnost
5e-3	50	84,55 %
5e-4	50	78,65 %
5e-5	50	75,35 %
1e-1	100	57,05 %
1e-2	100	84,90 %
1e-3	100	79,75 %
1e-4	100	76,70 %
1e-5	100	72,80 %
1e-5	300	75,45 %

3.4 Hlavní experiment

V tomto experimentu je provedeno 15 opakovaných experimentů se stále stejným nastavením hyperparametrů a stejnou sadou obrázků, která je promíchána v každé epoše během trénování (viz parametr `Shuffle` v [Tab. 3.4](#)). Těchto 15 opakování je uskutečněno pro každý barevný model: RGB, HSV, XYZ, LAB, YIQ a YCbCr. Celkově tedy 90 samostatných experimentů, které spočívají v trénování a testování CNN. Samostatné experimenty jsou dále v práci pojmenovány prostým pojmem *opakování*.

Nastavení `trainingOptions` je zobrazeno v [Tab. 3.4](#). Oproti inicializačnímu experimentu byl změněn počet epoch na 30.

Tab. 3.4 Nastavení hyperparametrů pro hlavní experiment

<code>solverName</code>	<code>sgdm</code>
<code>InitialLearnRate</code>	<code>0.01</code>
<code>MaxEpochs</code>	<code>30</code>
<code>MiniBatchSize</code>	<code>40</code>
<code>Shuffle</code>	<code>every-epoch</code>
<code>ValidationFrequency</code>	<code>150</code>

Výsledky experimentu je možno nalézt ve složce: `\prilohy\3_hlavni_experiment`. Výstupem experimentu je:

- matice záměn `cm.png` pro každé opakování,
- trénovací křivka `trainingProgress.png` pro každé opakování,
- přesnosti testování a doba trénování pro každé opakování v `data.mat`.

Pro přesnosti testování a doby trénování všech opakování jednotlivých barevných modelů byl vypočten průměr a medián. Pro přesnosti testování byla navíc vyhodnocena směrodatná odchylka (std). Shrnutí těchto výsledků ukazuje **Tab. 3.5**.

Tab. 3.5 Shrnutí výsledků hlavního experimentu

	RGB	HSV	XYZ	LAB	YIQ	YCbCr
průměr doby trénování	4.41 h	6.36 h	8.29 h	6.42 h	7.56 h	6.19 h
medián doby trénování	4.41 h	4.58 h	8.27 h	6.14 h	8.06 h	5.33 h
průměr přesnosti testování	92,43 %	83,37 %	90,75 %	89,35 %	90,64 %	90,48 %
medián přesnosti testování	92,50 %	83,25 %	90,60 %	89,35 %	90,80 %	90,60 %
std přesnosti testování	0,88 %	0,95 %	0,36 %	0,49 %	0,82 %	0,86 %

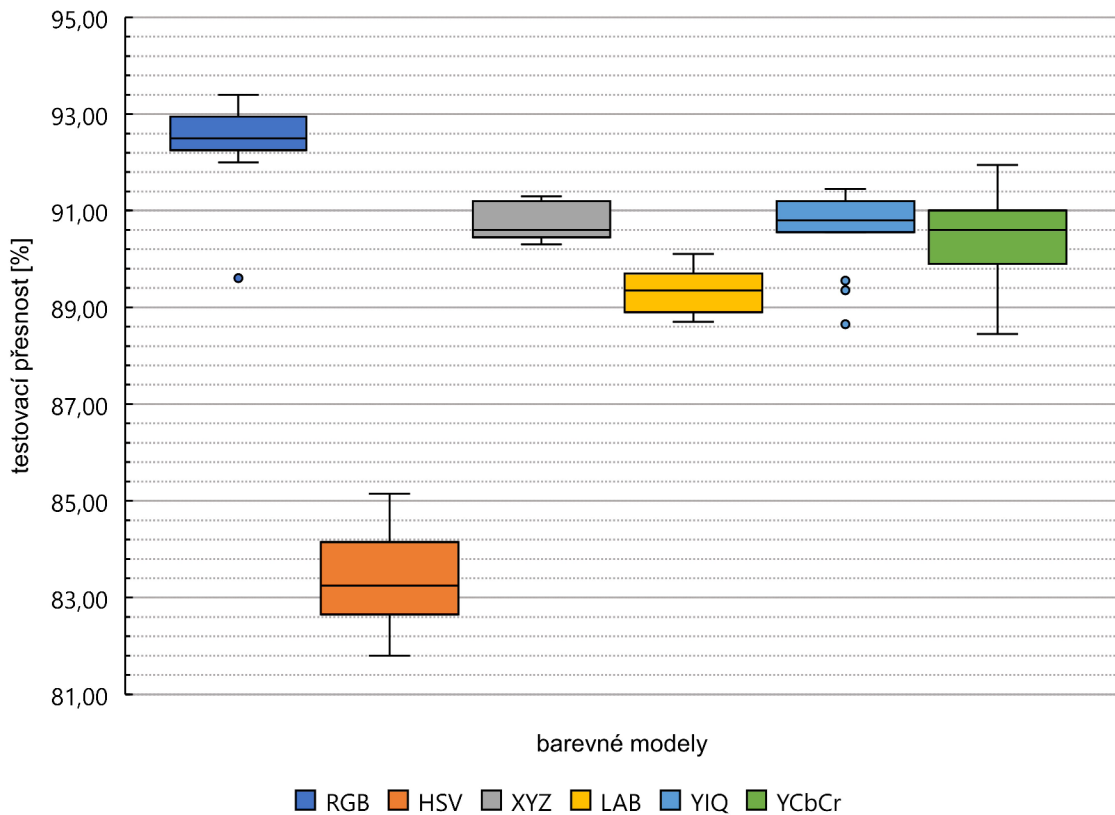
Z výše uvedené tabulky lze vidět, že nejhůře dopadl model HSV, což potvrzuje výsledek inicializačního experimentu (podkapitola 3.3). Barevné modely XYZ, LAB, YIQ a YCbCr si vedly jako u inicializačního experimentu velice podobně. Nejlépe dopadl model RGB. Tento poznatek potvrzuje také článek [9].

Zároveň lze podle průměrné doby trénování vidět, že transformace z modelu RGB do dalších barevných modelů se projevuje delší potřebnou dobou trénování. Tento poznatek sice nepotvrzuje model HSV u inicializačního experimentu, ale u tohoto hlavního experimentu je všech 15 opakování barevného modelu HSV déle trvajících. Trénování u 5. opakování modelu HSV trvá dokonce 13.26 h, tedy více než dvojnásobek průměru. Proto je vhodnější pro vyhodnocení doby trénování i testovací přesnosti použít medián. Obecně barevný model HSV je nejen v časech, ale i v přesnostech nejméně konzistentní, a proto je mu věnována hlubší pozornost v podkapitole 3.5. Pro zkrácení doby trénování je vhodné provést transformaci ještě před samotným trénováním.

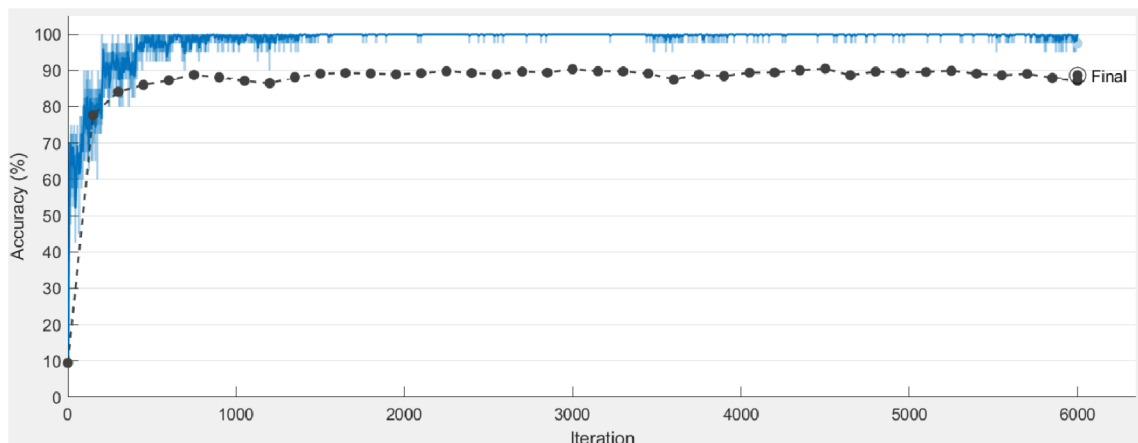
Z transformovaných modelů byl dle mediánu nejpresnější barevný model YIQ. Tento výsledek je možno vidět i v krabicovém grafu na **Obr. 3.2**. Na svislé ose je zobrazena testovací přesnost v rozmezí 81 % až 95 %. Jednotlivé boxy v krabicovém grafu představují barevné modely. Vodorovné čáry uprostřed boxů znázorňují medián. Barevná část boxu je tzv. interkvartilové rozpětí, zesponu ohraničená prvním kvantilem a shora třetím kvantilem. Mezi mediánem a prvním kvantilem se nachází 25 % dat. Taktéž mezi mediánem a třetím kvantilem. Mimo tuto barevnou část lze vidět tzv. vousy, jejichž koncové body představují minima/maxima přesností.

U modelu YIQ lze vidět tři odlehlé body (a jeden u modelu RGB). Jedná se o 5., 10. a 14. opakování. Tato nižší přesnost je způsobena náhlým rozkmitáním trénovací

křivky na konci trénování (6 000. iterace / 30. epocha), jak je to vidět na [Obr. 3.3](#). Jedna epocha je složena z 200 iterací jak u inicializačního, tak u hlavního experimentu.



Obr. 3.2 Porovnání jednotlivých barevných modelů z hlediska přesnosti testování



Obr. 3.3 Rozkmitání trénovací křivky v posledních iteracích má za následek snížení testovací přesnosti

Stejný jev je také případem nižší přesnosti u posledního opakování modelu RGB. Je vhodné vynechat tyto případy, aby neměly vliv na přesnost nebo stejně jako u trénovacího času vyhodnocovat medián, který je vůči odlehlým bodům robustní.

Soudě dle mediánu je v přesnosti testování u barevných modelů mimo model RGB nejlepší model YIQ. Rozdíly mezi barevnými modely XYZ, LAB, YIQ a YCbCr se pohybují v rozmezí 1,45 procentních bodu: 90,80 % u modelu YIQ vs. 89,35 % u modelu LAB v [Tab. 3.5](#).

Dále byly na základě experimentu vyhodnoceny matice záměn jednotlivých barevných modelů. Byl stanoven průměr a medián falešně negativních (false negatives, FN) výsledků klasifikace. Jsou to výsledky, u nichž CNN vyhodnotila skutečnou kategorii (true class) nesprávně. Výsledek pro průměry je možno vidět v [Tab. 3.6](#), pro mediány pak v [Tab. 3.7](#).

Tab. 3.6 Průměry falešně negativních odhadů pro všechny kategorie všech barevných modelů. Červeně jsou označeny nejhorší výsledky, zeleně nejlepší.

	RGB	HSV	XYZ	LAB	YIQ	YCbCr
baby back ribs	7,73	27,80	12,20	13,27	9,87	10,93
beef tartare	16,87	49,53	24,07	27,07	22,80	22,87
churros	17,67	34,60	20,60	22,40	20,40	21,73
donuts	22,07	44,33	28,13	31,80	29,93	31,93
omelette	11,87	29,93	11,93	19,53	15,13	14,33
onion rings	17,07	29,00	19,53	20,80	18,53	21,40
red velvet cake	16,27	30,40	19,13	22,33	17,87	17,53
spaghetti bolognese	8,20	20,60	9,87	13,33	10,20	9,80
sushi	13,93	30,00	16,93	18,80	20,33	18,53
tiramisu	19,67	36,47	22,67	23,60	22,20	21,40

Tab. 3.7 Mediány falešně negativních odhadů pro všechny kategorie všech barevných modelů. Červeně jsou označeny nejhorší výsledky, zeleně nejlepší.

	RGB	HSV	XYZ	LAB	YIQ	YCbCr
baby back ribs	6	28	11	14	9	9
beef tartare	16	44	24	27	21	21
churros	18	32	20	23	18	22
donuts	21	43	27	30	26	31
omelette	11	28	10	17	15	13
onion rings	17	27	19	19	17	20
red velvet cake	15	28	18	20	18	15
spaghetti bolognese	7	18	8	13	8	7
sushi	13	29	16	18	20	17
tiramisu	17	33	23	22	20	21

Podle tabulek výše je vidět, že rozdíl v počtu nejlepších kategorií, které jsou označeny zeleně, je mezi těmito dvěma tabulkami pouze u barevných modelů LAB a YIQ. Obecně lze říci, že kategorie spaghetti bolognese společně s baby back ribs byly zaměňovány nejméně. Nejhuře dopadla kategorie donuts, která byla nejčastěji zaměňována s kategoriemi churros, sushi a onion rings. Zde se s velkou pravděpodobností

jedná o tvarovou podobnost – kulaté donuty s dírou uprostřed, mohou být snadno zaměněny s cibulovými kroužky, jenž mají tvar velmi podobný. V některých případech mají dokonce i velmi podobnou barvu. Porovnání dvou tvarově a barevně podobných případů, jeden z kategorie churros, druhý z kategorie onion rings lze pozorovat na [Obr. 3.4](#).



(a) Churros



(b) Onion rings

Obr. 3.4 Porovnání dvou tvarově a barevně podobných obrázků ze dvou různých kategorií datové sady Food-101. Převzato z [3].

Vyhodnocení nejčastěji zaměňovaných kategorií jednotlivých barevných modelů – tedy maximální hodnota FN v maticích záměn – je následující: u barevných modelů RGB, XYZ, LAB, YIQ a YCbCr je to kategorie tiramisu a red velvet cake. Tyto dvě kategorie jsou tvarově velmi podobné, jedná se o dezerty, což je pravděpodobná příčina zvýšené hodnoty FN.

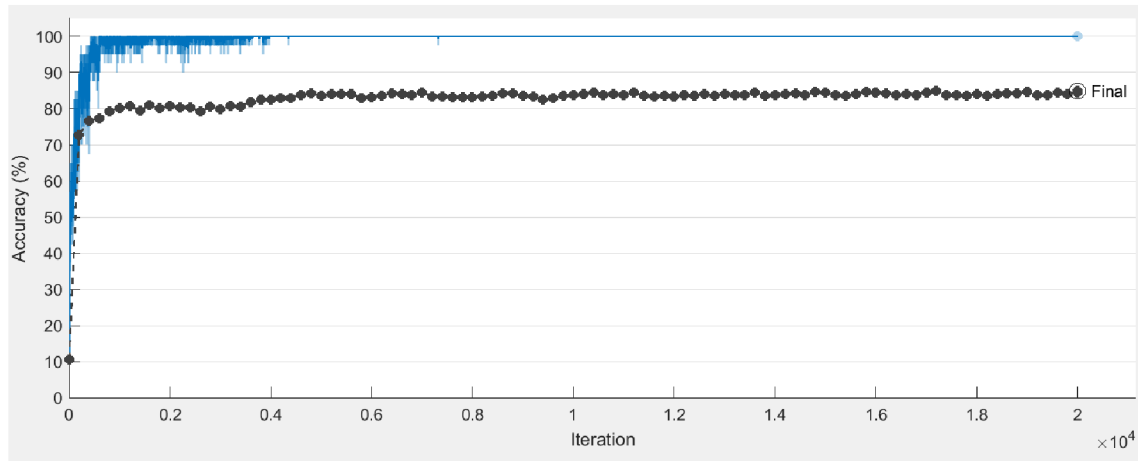
Aby se daly výsledky získané v této podkapitole zobecnit, stálo by za to provést experiment se stejnou datovou sadou, ale jinou architekturou CNN – například podle závěrů [9] architekturou Inception-V3, u níž dopadl barevný model LAB lépe než model RGB. Druhým doporučeným experimentem pro potvrzení chování architektury ResNet-101 je uskutečnění stejného experimentu, ale na dalších datových sadách.

3.5 Samostatné vyhodnocení modelu HSV

Model HSV se choval specificky. Nejen, že přesnosti testování a časy trénování jsou nejméně konzistentní, ale i celková přesnost zobrazená na [Obr. 3.2](#) je nejnižší. Cílem této podkapitoly je zjistit, proč tomu tak je.

Předně přesnost nebyla ovlivněna náhlým rozkmitáním trénovací křivky a tedy vznikem odlehlých bodů. Tuto skutečnost potvrzuje i inicializační experiment s nastá-

vením míry učení na $1e-2$, kde se trénovací křivka ustálila na 7000. iteraci a až do konce zůstala stabilní. Viz [Obr. 3.5](#). I u inicializačního experimentu dosáhl barevný model HSV nejnižší přesnosti testování mezi ostatními barevnými modely a nijak zásadně se neodlišuje od přesnosti testování u hlavního experimentu: 84,90 % vs. 83,25 %.

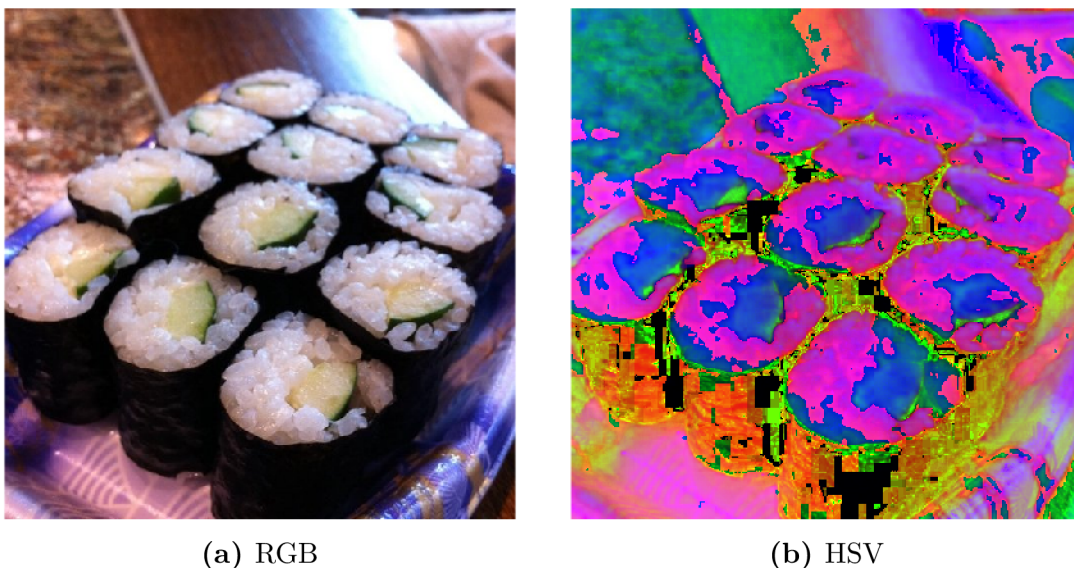


Obr. 3.5 Trénovací křivka u inicializačního experimentu barevného modelu HSV. Celkem 20 000 iterací, tzn. 100 epoch.

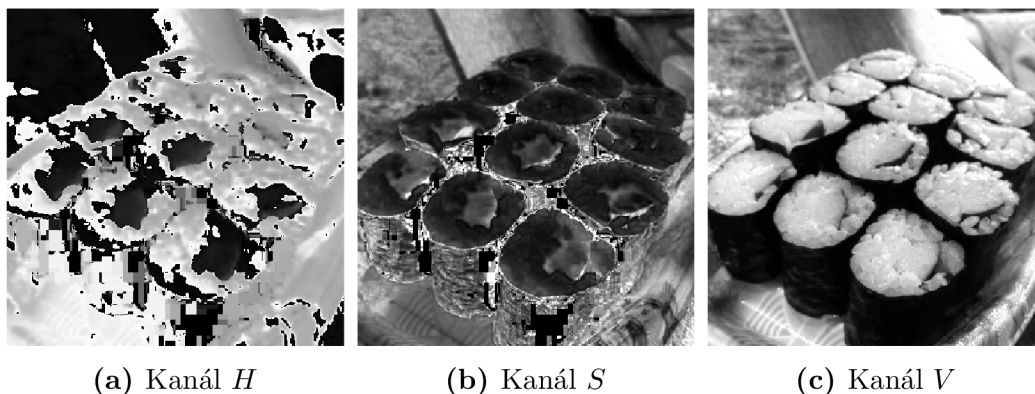
Další možností, jak prověřit sníženou přesnost, je vizuální porovnání obrázku transformovaného do barevného modelu HSV. Na [Obr. 3.6](#) je zobrazeno porovnání obrázku v RGB modelu a HSV modelu z kategorie sushi datové sady Food-101.

U barevného modelu HSV v [Obr. 3.6b](#) je vidět diskontinuita v barvách. V tmavých oblastech řas použitých pro zaobalení náplně sushi jsou ostré barevné přechody, hlavně černé a světle zelené barvy. Taktéž v náplni sushi je vidět ostré fialové mapy, přičemž dochází ke splnutí tvarů, tedy ke ztrátě jejich kvality. Tímto vysvětluje nižší přesnost testování také článek [5].

Vysvětlení této diskontinuity je nasnadě. Podle rovnic [2.1](#) v podkapitole [2.2.2](#) pro převod z barevného modelu RGB do modelu HSV je pro model HSV, konkrétně pro kanál H , kritický přechod mezi souřadnicemi v RGB modelu $[R, G, B] = [255, 0, 0]$ a $[R, G, B] = [255, 0, 1]$. V modelu RGB se tento rozdíl téměř neprojeví – červená zůstane červenou, ale v kanálu H modelu HSV hodnota skočí z jednoho extrému na druhý – z hodnoty 0 do hodnoty 1 (hodnoty HSV se pohybují v rozmezí $\langle 0, 1 \rangle$). Znázorněno ve stupních šedi je to skok z černé barvy do bílé. Tento případ je pozorovatelný v [Obr. 3.7a](#) a [Obr. 3.8](#) v prvním řádku. Druhý kritický přechod je na souřadnici $[R, G, B] = [0, 0, 0]$. Zde stačí navýšit jakoukoli složku o hodnotu 1, např. $[R, G, B] = [1, 0, 0]$, což se projeví v kanálu sytosti S opět skokem z jednoho extrému na druhý. Toto je vidět v [Obr. 3.7b](#) a [Obr. 3.8](#) v druhém řádku. Nevhodné jsou proto pro barevný model HSV červené nebo černé oblasti v obrázku RGB. V případě [Obr. 3.6](#) jsou to tmavé řasy používané jako obal pro náplň sushi.



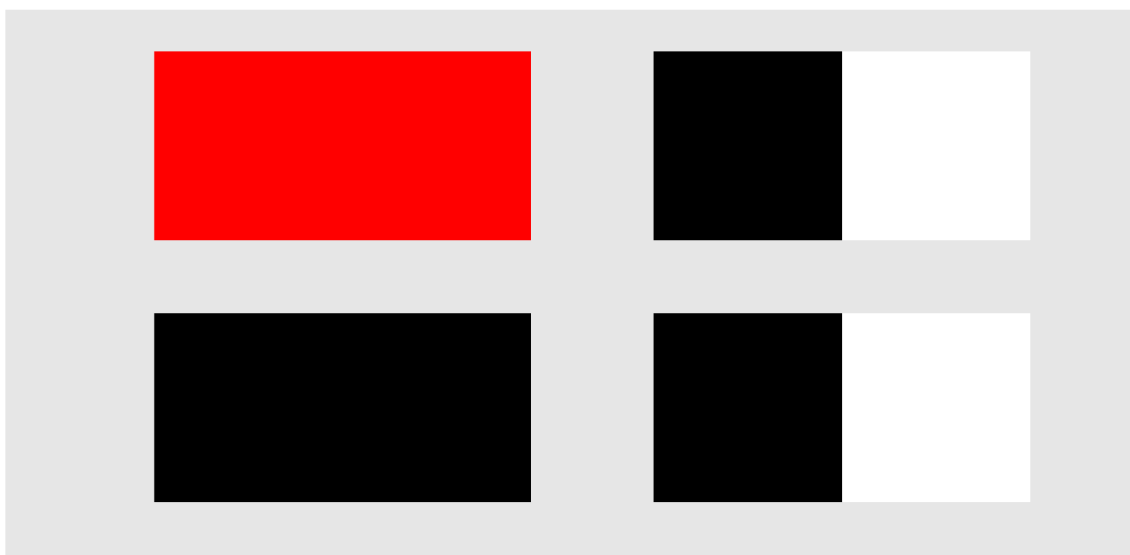
Obr. 3.6 Porovnání barevných modelů RGB a HSV obrázku z kategorie sushi datové sady Food-101. U modelu HSV lze pozorovat diskontinuitu barev. Převzato z [3].



Obr. 3.7 Porovnání jednotlivých kanálů barevného modelu HSV obrázku z kategorie sushi datové sady Food-101. Převzato z [3].

Z článků zmíněných v kapitole 2.3 model HSV obecně obsazoval nižší příčky v klasifikačních úlohách při použití CNN. Výhodným oproti barevnému modelu RGB je například v úloze detekce make-upu na obličeji [2]. Nicméně i pro tuto úlohu se ukázalo jako lepší použít barevný model CMYK. Autoři [5] model HSV doporučují pouze v případě, kdy je v klasifikační úloze důležitější barevná informace a méně důležitější jsou tvary.

Použití samotného barevného modelu HSV pro klasifikační úlohu CNN nelze doporučit. Lze jej ale použít v případech multi-barevných modelů. Články [13] a [35] s modelem HSV pracují v kombinaci s dalšími barevnými modely a dosahují díky tomu vyšších přesností.



Obr. 3.8 Znárodnění diskontinuity transformace z barevného modelu RGB do modelu HSV. Vlevo nahoře jsou zobrazeny dva pixely červené vedle sebe, jejichž rozdíl v barvě je nerozeznatelný, ale transformace do modelu HSV se projeví v kanále H ostrým přechodem z hodnoty 0 (černá) do hodnoty 1 (bílá), což je znázorněno vpravo nahoře. Dolní obrázek vlevo podobně ukazuje dva téměř totožné pixely černé v modelu RGB a vpravo pak jejich reprezentaci v kanálu S modelu HSV – opět lze vidět ostrý přechod při přidání jednoho bitu do jakékoli složky modelu RGB.

Kapitola 4

Závěr

Práce přinesla shrnutí aktuálních poznatků o vlivu barevných modelů na chování hlubokých konvolučních neuronových sítí. Nejvíce článků zabývajících se tímto tématem je z roku 2019. Publikované práce naznačují, že transformace vstupních dat do různých barevných modelů jsou potenciální cestou pro zvýšení přesnosti klasifikace obrazových dat s využitím konvolučních neuronových sítí.

Na základě rešerše byl proveden experiment, který testoval vliv šesti barevných modelů RGB, HSV, CIE 1931 XYZ, CIE 1976 L*a*b*, YIQ a YCbCr na chování architektury hluboké konvoluční neuronové sítě ResNet-101. Porovnání bylo realizováno na podmnožině datové sady Food-101, která byla tvořena deseti kategoriemi. Byla zároveň očištěna od duplicitních a tzv. parazitních obrázků, které se nacházely omylem v dané kategorii. Smazané obrázky byly doplněny z jiné, ale velmi podobné datové sady UMPC Food-101. S těmito datovými sadami nebyl doposud proveden experiment tohoto typu, což lze pokládat za další přínos této práce. Pro každý barevný model bylo provedeno 15 totožných opakování pro statistické vyhodnocení.

Z experimentu vyplývá, že barevný model HSV je pro zvolenou datovou sadu a architekturu nejméně vhodný. Celková přesnost tohoto modelu byla nejnižší, pravděpodobným důvodem je jeho diskontinuita vznikající při transformaci z modelu RGB. Dva sobě velmi blízké odstíny červené v modelu RGB, které jsou pohledem nerozeznatelné, se v modelu HSV projevují skokově. Dochází tak ke ztrátě tvarů. Použití samotného modelu HSV tudíž autor nedoporučuje.

Barevný model RGB se jeví jako nejvhodnější pro danou úlohu. Pozitivní vliv transformace do jiných barevných modelů nebyl prokázán. Pro získání obecně platných závěrů bude ale potřeba uskutečnit navržený experiment jednak s totožnou datovou sadou, ale jinou architekturou CNN (např. Inception-V3), jednak s totožnou architekturou, ale na jiných datových sadách.

Literatura

- [1] BASAVESWARA, S. K. *CNN Architectures, a Deep-dive* [online]. Towards Data Science, 27. srpna 2019 [cit. 2020-06-17]. Dostupné z: <https://bit.ly/3dsiBhL>.
- [2] BERTACCHI, M. G. a SILVEIRA, I. F. Facial Makeup Detection using the CMYK Color Model and Convolutional Neural Networks. In: *2019 XV Workshop de Visão Computacional (WVC)*. IEEE, 2019, s. 54–60. ISBN 978-1-7281-5337-7.
- [3] BOSSARD, L., GUILLAUMIN, M. a VAN GOOL, L. Food-101 – Mining Discriminative Components with Random Forests. In: *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014, s. 446–461. ISBN 978-3-319-10598-7.
- [4] BROWNLEE, J. A Gentle Introduction to Computer Vision. *Machine Learning Mastery* [online]. Machine Learning Mastery, 19. března 2019. Aktualizováno 5. 7. 2019 [cit. 2020-06-17]. Dostupné z: <https://bit.ly/3hMmCkG>.
- [5] BUHRMESTER, V. et al. Evaluating the Impact of Color Information in Deep Neural Networks. In: *Pattern Recognition and Image Analysis*. Cham: Springer International Publishing, 2019, s. 302–316. ISBN 978-3-030-31331-9.
- [6] *Convolutional Neural Network* [online]. The MathWorks [cit. 2020-06-17]. Dostupné z: <https://bit.ly/2zRY8p0>. Path: Solutions; Deep Learning; What are Convolutional Neural Networks?
- [7] DEMBSKI, J. a SZYMAŃSKI, J. Bees Detection on Images: Study of Different Color Models for Neural Networks. In: *Distributed Computing and Internet Technology*. Cham: Springer International Publishing, 2019, s. 295–308. ISBN 978-3-030-05366-6.
- [8] DERTAT, A. Convolutional Neural Networks. *Applied Deep Learning* [online]. Towards Data Science, 8. listopadu 2017 [cit. 2020-06-17]. Dostupné z: <https://bit.ly/2Zb0vub>.

- [9] DIAZ CELY, J. et al. The Effect of Color Channel Representations on the Transferability of Convolutional Neural Networks. In: *Advances in Computer Vision*. Cham: Springer International Publishing, 2020, s. 27–38. ISBN 978-3-030-17794-2.
- [10] DOS SANTOS, T. J. et al. On the Influence of the Color Model for Image Boundary Detection Algorithms based on Convolutional Neural Networks. In: *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019, s. 1–8. ISBN 978-1-7281-1985-4.
- [11] GONZALEZ, R. C. a WOODS, R. E. *Digital Image Processing*. Pearson, 2018. ISBN 978-1-292-22304-9.
- [12] GOODFELLOW, I., BENGIO, Y. a COURVILLE, A. *Deep Learning*. MIT Press, 2016. ISBN 978-0-262-03561-3. <http://www.deeplearningbook.org>.
- [13] GOWDA, S. N. a YUAN, C. ColorNet: Investigating the Importance of Color Spaces for Image Classification. In: *Computer Vision – ACCV 2018*. Cham: Springer International Publishing, 2019, s. 581–596. ISBN 978-3-030-20869-1.
- [14] HE, K. et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2015, s. 1026–1034. ISBN 978-1-4673-8391-2.
- [15] HE, K. et al. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2016, s. 770–778. ISBN 978-1-4673-8851-1.
- [16] HUANG, G. et al. Densely Connected Convolutional Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, s. 2261–2269. ISBN 978-1-5386-0457-1.
- [17] KOHOUT, V. Historie a elementární základy teorie barev III. *Školská fyzika*. Fraus. Červen 2013, X, č. 2, s. 41–48. Fyzika kolem nás. ISSN 1211-1511.
- [18] KOLESNIKOV, A. et al. Big Transfer (BiT): General Visual Representation Learning. *ArXiv*. 2019, abs/1912.11370.
- [19] KRIZHEVSKY, A., SUTSKEVER, I. a HINTON, G. ImageNet Classification with Deep Convolutional Neural Networks. In: *Advances in Neural Information Processing Systems 25*. Curran Associates, 2012, s. 1097–1105.
- [20] LECUN, Y., KAVUKCUOGLU, K. a FARABET, C. Convolutional networks and applications in vision. In: *Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. IEEE, 2010, s. 253–256. ISBN 978-1-4244-5308-5.

- [21] LINDBLOOM, B. J. *RGB to XYZ* [online]. Bruce Justin Lindbloom, 8. dubna 2017 [cit. 2020-06-17]. Dostupné z: <https://bit.ly/2Cwd2ST>.
- [22] MEINEL, C. a SACK, H. *Digital Communication*. Springer-Verlag Berlin Heidelberg, 2014. ISBN 978-3-642-54330-2.
- [23] NG, C., TAY, Y. a GOI, B. Comparing Image Representations for Training a Convolutional Neural Network to Classify Gender. In: *2013 1st International Conference on Artificial Intelligence, Modelling and Simulation*. IEEE, 2013, s. 29–33. ISBN 978-1-4799-3251-1.
- [24] OPENAI et al. Dota 2 with Large Scale Deep Reinforcement Learning. *ArXiv*. 2019, abs/1912.06680.
- [25] PAR. *CIE 1931 color space chromaticity diagram visualizing the sRGB gamut* [online]. Wikimedia Commons, 25. května 2005. Aktualizováno 26. 5. 2008 [cit. 2020-06-17]. Dostupné z: <https://bit.ly/2YrHIxh>.
- [26] PINGEL, J. *Practical Deep Learning Examples with MATLAB* [online]. The MathWorks, 2020 [cit. 2020-06-17]. Dostupné z: <https://bit.ly/3emD3SM>.
- [27] POYNTON, C. A. *Frequently Asked Questions about Color* [online]. Březen 1997 [cit. 2020-06-17]. Dostupné z: <https://bit.ly/2Z1ALSm>.
- [28] *Pretrained Deep Neural Networks* [online]. 2020 [cit. 2020-06-17]. Dostupné z: <https://bit.ly/3dn8bQJ>. Path: Support; Documentation; Deep Learning Toolbox; Deep Learning with Images.
- [29] SILVER, D. et al. Mastering the game of Go with deep neural networks and tree search. *Nature*. 2016, sv. 529, č. 7587, s. 484–489. ISSN 1476-4687.
- [30] SIMONYAN, K. a ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv*. 2015, abs/1409.1556.
- [31] SMITH, A. R. Color Gamut Transform Pairs. In: *Proceedings of the 5th Annual Conference on Computer Graphics and Interactive Techniques*. New York: Association for Computing Machinery, 1978, s. 12–19. ISBN 978-1-4503-7908-3.
- [32] SZEGEDY, C. et al. Going Deeper with Convolutions. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015, s. 1–9. ISBN 978-1-4673-6964-0.
- [33] SZEGEDY, C. et al. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. *ArXiv*. 2017, abs/1602.07261.

- [34] SZELISKI, R. *Computer Vision: Algorithms and Applications*. 1. vyd. Springer-Verlag London, 2010. ISBN 978-1-84882-934-3.
- [35] TAN, J. a LI, N. Ensemble Learning Based Multi-Color Space in Convolutional Neural Network. In: *2019 Chinese Control Conference (CCC)*. IEEE, 2019, s. 7924–7927. ISBN 978-1-7281-2329-5.
- [36] TOUVRON, H. et al. Fixing the train-test resolution discrepancy: FixEfficientNet. *ArXiv*. 2020, abs/2003.08237.
- [37] *Understanding Color Spaces and Color Space Conversion* [online]. [cit. 2020-06-17]. Dostupné z: <https://bit.ly/381BnMd>. Path: Support; Documentation; Image Processing Toolbox; Import, Export, and Conversion; Color.
- [38] WANG, X. et al. Recipe recognition with large multimodal food dataset. In: *2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. 2015, s. 1–6. ISBN 978-1-4799-7079-7.
- [39] XIE, Q. et al. Self-training with Noisy Student improves ImageNet classification. *ArXiv*. 2019, abs/1911.04252.