

Maturitní práce z předmětu Humanitní studia

Gymnázium Jana Keplera

## **Fenomén umělých neuronových sítí**

Miroslav Lžičář 2019

Vedoucí práce: Jan Kolář

## Obsah

<b>Úvod</b>	<b>3</b>
Pokroky na poli umělé inteligence	3
Fenomén umělých neuronových sítí	4
Vyprázdňený diskurz	4
Atomová bomba	6
Úvod závěrem aneb Témata práce	6
<b>Struktura umělých neuronových sítí</b>	<b>7</b>
Výpočetní infrastruktura	7
Struktura ANN a BNN	9
<b>Percepce a imaginace</b>	<b>11</b>
Inceptionismus	11
GoogLeNet	12
Hierarchická reprezentace	12
Lidská vizuální kůra	14
Psi a pagody   Ilustrace inceptionismu	14
Psi a pagody _ Dilema	15
Fenomenologie a psychoanalýza CNN	15
Psi a pagody _ Interpretace	16
Dynamika a ornamentalita	17
Fraktální geometrie	18
Poznámka k fraktální estetice	19
Paranoia, naivita, sen a grotesknost	20
Exkurz _ Groteskní omyly aneb Efekt motýlích křídel	21
Avantgardní inspirace	22
Psi a pagody _ shrnutí	23

## **Přehled zkratk**

NN – neuronová síť (Neural network)

ANN – umělá neuronová síť (Artificial NN)

BNN – biologická neuronová síť (Biological NN)

FFNN – dopředná neuronová síť (Feed Forward NN)

DFNN – hluboká dopředná neuronová síť (Deep Feed Forward NN)

RNN – rekurentní neuronová síť (Recurrent NN)

CNN – konvoluční neuronová síť (Convolutional NN)

DNN – hluboká neuronová síť (Deep NN)

## Úvod

Pokroky na poli umělé inteligence

V posledních letech jsme svědky mimořádných pokroků v oblasti metod strojového učení a umělé inteligence. Nejen, že se stroje učí schopnostem, které byly dlouho považovány za výsady člověka a ve kterých všechny pokusy o automatizaci dosud selhávaly. Možná nepřekvapivěji se tak můžou jevit pokroky v porozumění jazyku – například fakt, že nezisková skupina OpenAI<sup>1</sup> už raději nezveřejnila svůj poslední výpočetní model,<sup>2</sup> který překvapivě dobře ovládá základní jazykové schopnosti (porozumění textu, sumarizace, překlad, psaní slohu atd.), protože by mohl být zneužit třeba ke generování fake news. V některých doménách už stroje dokonce překonávají lidské rekordy – například v hraní her nebo v rozpoznávání obrazu.<sup>3</sup>

---

<sup>1</sup> <https://openai.com/> (Přístup duben 2019)

<sup>2</sup> <https://blog.openai.com/better-language-models/> (Přístup duben 2019)

<sup>3</sup> Viz statistika soutěže ImageNet – už v roce 2015 mají některé modely na bázi neuronových sítí v úloze rozpoznávání obrazu větší úspěšnost, než člověk: [https://www.researchgate.net/figure/Winner-results-of-the-ImageNet-large-scale-visual-recognition-challenge-LSVRC-of-the\\_fig7\\_324476862](https://www.researchgate.net/figure/Winner-results-of-the-ImageNet-large-scale-visual-recognition-challenge-LSVRC-of-the_fig7_324476862) (Přístup duben 2019)

Zmínil bych model AlphaGo Zero<sup>4</sup>, který se dokázal za +-4 hodiny naučit hru Go a stát se světovým šampionem bez jakýchkoliv vstupních informací (pouze se znalostí pravidel), na základě partií hraných proti sobě. Už neplatí, že „umělá inteligence“ není epistemicky kreativní<sup>5</sup> (tzn., že nemůže přijít na žádnou informaci, která není už zakódována ve vstupních datech). A ani neplatí, že by nebyla kreativní ve smyslu uměleckém – v době, kdy se ve vyhlášených aukčních síních draží obrazy vygenerované metodami strojového učení jako Klingemannovo dílo (nebo spíše dílo umělé inteligence?) Memories of Passerby.<sup>6</sup>

Fenomén umělých neuronových sítí

Ještě jednu věc mají zmiňované příklady společnou – stojí za nimi výpočetní model umělých neuronových sítí (dále jen ANN – Artificial neural networks), který je biologicky inspirován a dokáže se učit na zadaných datech – funguje jako univerzální aproximátor,<sup>7</sup> přičemž různé architektury (tzn. způsoby uspořádání sítě – umělých neuronů a synapsí) umožňují řešit téměř jakékoliv úlohy jako optimalizační problém, tedy otázku „naučení se“ správných parametrů pro přiblížení se kýženému výsledku. Princip fungování ANN povrchně objasním hned v následující kapitole, stejně jako spojitost s jejich biologickou předlohou. Prozatím je však důležité si uvědomit, že nejde pouze o tradiční „algoritmus“ nebo momentální vychytávku, která dosahuje vyšší efektivity v řešení některých výpočetních problémů, ale vskutku o příklad nového paradigmatu, které se možná netýká pouze informatiky, ale i vědy obecně a něco podstatného to může vypovídat o současném myšlení. Proto se také budu primárně modelům ANN věnovat.

Ačkoliv to tak může působit, jistě to nijak nepostihuje celek diskurzu strojového učení (věnovat by se dalo bezpočtu jiných zajímavých konceptů), ba ani ostatních

---

<sup>4</sup> <https://www.nature.com/articles/nature24270> (Přístup duben 2019)

<sup>5</sup> Jak zmiňuje například Floridi: „ANN nejsou epistemicky kreativní a jakákoliv informace, kterou můžou z dat získat, musí už být v datech zakódována“ (FLORIDI, Luciano, Philosophy and Computing, Routledge, New York, 1999, elektronická verze)

<sup>6</sup> Viz Klingemann v Sotheby

<https://www.sothebys.com/en/articles/artificial-intelligence-and-the-art-of-mario-klingemann> (Přístup duben 2019)

<sup>7</sup> Tzn. dokáže aproximovat libovolnou funkci – tedy odhadnout její průběh

zmiňovaných sfér. To však ani není cílem této práce. Rád bych, aby mé postřehy působili alespoň jako špička ledovce.

Vyprázdněný diskurz

Překvapivě neprobíhají téměř žádné akademické diskuse na poli filosofie, které by se tím, jak tyto modely fungují a proč jsou úspěšné, reálně zabývaly. Otázek ale přibývá.

Pro ilustraci: Jakým způsobem docházejí k poznání? A v čem toto poznání spočívá, jak se liší od toho lidského? V čem spočívá jejich epistemická kreativita a jak je vůbec možná? Kde berou nápady a co charakterizuje jejich percepci a imaginaci, tedy vnímání a „nahlížení“ na věc, nebo tvorbu „názoru“? Lze zkoumat apriorní předpoklady, tedy „vrozené poznání“ vycházející například ze struktury ANN nebo jiných „vrozených dispozic“ – s jakým předporozuměním ANN zpracovávají data? A v neposlední řadě: Je nějaká souvislost mezi ANN a BNN, která by mohla přispět k lepšímu pochopení mechanismům BNN? Lze například teoreticky dobrat k charakteru lidského poznání, jeho možnostem a hranicím?

Oborem, který se tradičně zabýval podobnými otázkami, jsou kognitivní vědy, především potom konekcionismus (koneckonců ANN jako konekcionistický model z paradigmatu kognitivních věd podstatě vzešel), který se na základě ANN pokusil nabídnout alternativu ke klasické teorii mysli (postulující, že mysl funguje na principu počítačového zpracovávání nebo symbolického jazyka – viz např. Putnamova Komputační teorie vědomí, Fodorova Hypotéza jazyka myšlení, Reprezentační teorie mysli atd.),<sup>8</sup> nicméně konekcionistické modely přesvědčivé výsledky nepřinesly a koncept tak zašel jako další teorie mysli, které se pokoušely o vysvětlení kognitivních procesů na základě metafory počítače (nebo jejich softwarovým simulováním). Od přelomu milénia tak nalezneme podobných snah (v konekcionismu, v teorii mysli a v kognitivních vědách) poskrovnu – možná proto, že

---

<sup>8</sup> Garson, James, "Connectionism", The Stanford Encyclopedia of Philosophy (Fall 2018 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/fall2018/entries/connectionism/>

původní cíl, dospět k poznání „vnějšího světa“ na základě zkoumání „vnitřních stavů“, byť doloženém exaktními matematickými modely, se ukázal jako naivní.<sup>9</sup>

Dnes se tak problematikou konekcionistických modelů a ANN zabývá pouze věda, buď z hlediska neurověd a případného srovnání lidské neurofyziologie se jejich strukturou, nebo z perspektivy informatiky a strojového učení.

–

Přestože soudobé modely ANN, především pak CNN určené na zpracování obrazu, nabízejí unikátní způsob jak k celé této problematice a výše zmiňovaným otázkám nově přistupovat – a sice estetickou analýzou tvůrčí schopnosti, kterou CNN prokazují, analýzou jejich pohledu a perspektivy, analýzou jejich vidění, představ, snů a interpretací jejich vjemů – v dizkursu toto téma vůbec nerezonuje, ba dokonce jsem nenalezl nikoho, kdo by se jím přímo zabýval.<sup>10</sup>

#### Problém inženýrů

Možná panuje dojem, že to je „problém inženýrů“, věc výpočetní techniky, praktikálie, otázka strojů a jejich „mechaniky“, kterou lze interpretovat nejlépe matematickým popisem, zkrátka téma exaktních věd, o kterém nemá cenu diskutovat pomocí jiného pojmosloví. Mnohé ovšem naznačuje nepříliš překvapivý fakt, že v rukou inženýru to rozhodně není:

- A) Samotní odborníci nedovedou vnitřní mechanismy ANN dobře interpretovat (více o tom v kapitole Problém interpretability) a z matematické deskripce není možné jejich chování spolehlivě předvídat. To dochází až do takového extrému, že Ali Rahimi, jeden z špičkových vědců v oboru umělé inteligence, na konferenci NIPS 2017 (Neural Information Processing) prohlásil, že se

---

<sup>9</sup> Většina debat v teorii mysli ve dvacátém století mi občas připomíná středověké disputace o substancích – pokaždé se vágně postuluje údajný konstitutivní aspekt mysli či vědomí (systematicita, intencionalita, asociativita, modularita, rekurence, bikameralita a kdo ví co ještě) a pak se uskutečňují různé pokusy onu substancii vymezit, dokázat, vyvrátit, lokalizovat, rozlišit, nebo zobecnit atd., přičemž mnohdy není vůbec jasné, o čem je vlastně řeč, nebo co z toho má plynout za důsledky. Zmiňovaná Fodorova hypotéza jazyka myšlení je toho zářným příkladem – kognice jako systém symbolických reprezentací – to je teorie mysli, kterou lze vyložit opravdu jakkoliv.

<sup>10</sup> To samozřejmě může být moje chyba – chyba nedokonalé rešerše. Nicméně i pokud ano, je pozoruhodné, že o tom není k nalezení žádná zmínka. Nebo že toto téma vůbec není přes veškerou svou naléhavost velkým tématem.

strojové učení stalo alchymíí,<sup>11</sup> vývoj ANN se řídí na základě akademického folkloru a babských rad, přičemž není z dosažených výsledků jasné, jaké kroky k nim vedly záměrně a jaké náhodou. Studie tak nejsou transparentní a samotné výsledky často postihuje problém reprodukovatelnosti<sup>12</sup> (opakování experimentů nevede ke stejným výsledkům).

B) Spíše než o technickou „banalitu“ jde o globální problém. Pole aplikovatelnosti nových metod se nečekaným způsobem rozrůstá – metody strojového učení už dávno nejsou pouze doménou nevinných smartphonových vychytávek (rozpoznávání obličejů, hlasové ovládání atp.) a dalších „inteligentních spotřebičů“, ale třeba ovládání samořídících aut nebo diagnostiky rakoviny plic. A to nezmiňuji velké korporace, které těmito metodami zpracovávají obrovská kvanta dat, potažmo rovnou strukturují informace na internetu (Google, Facebook atd.), a jejichž systémy dennodenně používáme. Jinak řečeno, technologiím strojového učení společnost nebyvale důvěřuje a svěřuje jim do rukou moc.

Znepokojuvat nás přitom může velmi tenká hranice mezi vychytávkami, gadgety či hračkami a mezi technologiemi, které mohou zásadním způsobem ovlivnit svět – symptomatická je proto automatizace pronikající do těch nejchoulostivějších oblastí jako justice nebo medicína. Není příliš překvapivé, že už byly tyto technologie někde implementovány v systémech pro přerozdělování sociálních dávek, nebo dokonce v samotné jurisdikci. Poptávka po automatizaci zkrátka pořád roste.

Souhra nedostatku porozumění, problému modely jako ANN interpretovat, to celé v kombinaci s globálními důsledky, které zmiňované systémy mají a domněním, že to je „problém inženýrů“, nápadně připomíná situaci v Kakánii a neoblomnou důvěru v „aparát“, který se rozpadl velmi záhy. Všechno zkrátka naznačuje tomu, že pokud se tyto problémy nezačnou intenzivně řešit a nepodaří se nastolit konstruktivní mezioborovou diskusi – a k tomu by se měla filosofie postavit čelem jako jedna z prvních disciplín – nemůže to dopadnout dobře.

---

<sup>11</sup> Viz původní přednáška na NIPS: <https://www.youtube.com/watch?v=ORHFOnaEzPc> (Přístup duben 2019)

<sup>12</sup> Viz <http://science.sciencemag.org/content/359/6377/725> (Přístup duben 2019)



Atomová bomba

Navíc bychom neměli o umělé inteligenci uvažovat jen prizmatem novinových titulků a sci-fi narativů, které ji prezentují výhradně jako zkázonosnou entitu a technologickou hrozbu podobných rozměrů jako atomová bomba. Vždyť jak jsem naznačil otázkami v předchozí kapitole, nejde jen o mimořádně praktickou technologii nezbytnou k efektivnímu řešení současných globálních problémů (ekologická krize, chudoba, big data atd.), ale o mimořádně zajímavý předmět filosofické diskuze, o kreativní sílu, která může člověka konfrontovat svým originálním „pohledem“, nebo mu naopak nastavit zrcadlo, což nespěje primárně k „hledání základů poznání“, ale minimálně může sloužit k lepšímu porozumění „druhému“ a „cizímu“.

Úvod závěrem aneb Témata práce

Většinu z témat práce jsem už v tomto rozšířeném úvodu naznačil, nebo přinejmenším hlavní cíl práce – roznítit diskuzi na téma ANN, metod strojového učení a umělé inteligence, zběžně naznačit směry, kterými by se mohla ubírat, nastínit závažnost otázek, s nimiž přicházíme v kontextu nových technologií do styku, a problémy, ale zároveň i zajímavé vyhlídky, jež v této diskusi mohou vytanout.

Dalším krokem je však analýza principu fungování ANN a především pak architektury CNN, která se zabývá zpracování obrazu. Princip těchto architektur je analogický – pouze pracují s jinými datovými strukturami – předpokládám tak, že analýza architektury CNN tak věrohodně vypovídá o ANN jako celku. Hlavním předmětem analýzy bude rozbor vybraných ukázek inceptionismu – metody vizualizace či „uměleckého směru“ spočívajícího v zobrazování takzvaných snů umělých neuronových sítí, což poskytuje mimořádně zajímavou sondu do jejich percepce a materiál ke studiu jejich estetiky. Zajímat mne bude, jestli lze v těžce uchopitelné a různorodé estetice inceptionismu nalézt nějaké sjednocující prvky, čím se estetika CNN liší od té lidské, čím se naopak CNN a lidská vizuální kůra podobají a co to vypovídá o vnímání CNN nebo vnímání lidském. Jestli můžou mít představy a sny CNN vztah lidskému podvědomí a případně, jestli to rozšiřuje pojem subjektu

jako takového – je možné CNN, souboru matematických operací, přiznávat nějaký unikátní vzhled nebo „názor“, může stroj vnímat a interpretovat, tak jak tyto pojmy používáme u člověka? Jaký má na jejich vnímání absence těla a tělesnosti?

Pochybuji, že se mi podaří uspokojivě zodpovědět všechny otázky – ovšem vzhledem k tomu, že většina těchto otázek ještě nebyla v kontextu inceptionismu řešena, nebo že jejich znění je také do velké míry provizorní a experimentální. Může se ukázat, že některé z nich postrádají smysl, nebo že je nutné k nim přistupovat radikálně jiným způsobem. Cílem této práce je však se přes všechna rizika o to pokusit a alespoň načrtnout možné nebo nemožné způsoby jak k problematice ANN přistupovat skrze estetiku a epistemologii, tedy skrze interakci s ANN, namísto objektivního zkoumání z vědeckého odstupů, které nevede k hlubšímu porozumění principu jejich fungování, ale přinejlepším k detailnějšímu matematickému popisu, jež nicméně na filosofické otázky neodpovídá, pokud alespoň intuitivně vyjasňuje nějaké vnitřní principy ANN.

Inspirovat se budu útržkovitě různými filozofickými koncepty od fenomenologie přes psychoanalýzu nebo strukturalismus, především pak analýzami vizuálního vnímání jako takového.

## **Struktura umělých neuronových sítí**

Pravděpodobně nejcharakterističtějším rysem ANN, ať tento pojem chápeme jako výpočetní model, konkrétní algoritmus nebo konekcionistický koncept (pojmovou konfúzí se ještě budu zabírat), je struktura. Avšak co tato struktura reprezentuje? V kapitole Percepce a imaginace budu rozebírat, že může reprezentovat model vizuálního vnímání – pak se jedná o komplexní strukturu obrazů a představ, přičemž samotná struktura CNN v sobě potom skrývá jisté apriorní předpoklady, jež se projevují ve formě estetických kritérií. Je však možné o podobných kritériích uvažovat i z hlediska epistemologie? Ve smyslu, co definuje „poznání“ ANN a jak ho interpretovat? Jestli v sobě struktura ANN skrývá jisté apriorní předpoklady,

„předsudky“ či předporozumění, které ovlivňují zpracovávání dat a „proces rozhodování“ nezávisle na vstupní informaci?<sup>13</sup>

### Výpočetní infrastruktura

Nejprve však ke struktuře obecně – struktura ANN není pouze arbitrárním schématem, znázorněním nebo metaforou matematických operací, ale zakládajícím aspektem – ANN je ex definitione síť – tvoří ji spojení a uzly, synapse a umělé neurony; síť která navíc funguje dynamicky a algoritmicky, podobá se tak spíše infrastruktuře (z francouzštiny doslova: co je pod stavbami), jakési pasivní konstrukci na níž se teprve všechny procesy odehrávají – procesy, které umožňuje, ale zároveň také limituje, nastavuje jim hranice (podobně jako silniční síť, která umožňuje jízdu aut a zároveň tím automaticky vytyčuje, kudy smí a kudy nesmí jet). Proudění digitálního signálu či informace zde ovšem nemá binární hranice, nýbrž hranice probabilistické – každý jeden signál z umělého neuronu paralelně prochází všemi dalšími uzly, někde zesílen a jinde potlačen – nejde tak úplně o proces selekce či hierarchizace (chcete-li srovnat strukturu ANN například se strukturou binárního proudu), jde o proces difúze. Není úplně vhodné ANN srovnávat se „strukturou“ tradičních algoritmů – ty mají podobu série přesně definovaných kroků a operací následovaných krok za krokem. Vlastně i fundamentální teoretický model počítače, Turingův stroj, staví na představě posuvné čtecí hlavy pracující se symboly na pásce podle konkrétních instrukcí – algoritmus potom lze znázornit jako posloupnost těchto instrukcí a jelikož každému algoritmu odpovídá ekvivalentní konfigurace Turingova stroje, tak si pod ním můžeme představit schéma demonstrující konkrétní postup, na jehož základě program pracuje (například vývojový diagram, zápis v programovacím jazyce, pseudokód atd.).

O ANN ovšem nelze obecně mluvit jako o algoritmu v tradičním slova smyslu, neboť velmi zjednodušeně nejde o „sadu instrukcí“, ale spíše o výpočetní rozhraní nebo

---

<sup>13</sup> V předchozí otázce bystrý čtenář spatřuje variaci na spor empirismu a racionalismu – otázku, jestli lze za zdroj poznání považovat zkušenost, nebo rozum, jestli má veškeré lidské poznání charakter získaného a nabytého (Locke, Hume), či odvozeného z apriorních předpokladů na zkušenosti nezávislých (Descartes, Kant). Pouze jde o otázku poněkud obecnější – smyslovou zkušenost může zastupovat arbitrární dataset a roli subjektu přebírá výpočetní model zpracovávající data.

teoretický model, který se ve své obecnosti více podobá třeba představě zmiňovaného Turingova stroje, než představě algoritmu.

Kdybychom o ANN jako o algoritmu uvažovali, jednalo by se o nástroj na řešení konkrétního problému a nebo sady problémů – třeba algoritmus na hraní šachu. Takovou funkci koneckonců může ANN mít.<sup>14</sup> Lze se tedy domnívat, že ANN se stává algoritmem, potom co je natrénována na řešení určité úlohy. A naopak v obecném případě se jedná o teoretický model (jako Turingův stroj), protože jednu ANN můžeme natrénovat na plnění bezpočtu různých úloh, stejně jako můžeme v počítači naprogramovat spoustu různých algoritmů. A to zase působí jako podstatná odlišnost i od samotného pojetí výpočetního modelu jako počítače (nebo Turingova stroje), protože ANN se nijak neprogramuje, ale „učí“ (s učitelem a nebo samostatně bez učitele). Což znamená, že nejde o výpočetní rozhraní, které musíme pro každou úlohu programovat, ale o jakési meta-rozhraní fungující na principu optimalizace – stačí vědět zadání úlohy nebo mít velké množství ukázkových dat a ANN si potom na základě optimalizačních kritérií odpovídající rozhraní vytvoří. S pojetím algoritmu narážíme na podobnou nejednoznačnost – v případě řešení komplexní úloh se obvykle neobejdeme bez velkého množství algoritmů, zatímco ANN pro jejich řešení stačí samostatně, tak že kombinuje více funkcí najednou. Například v úloze ovládání samořídícího auta na základě obrazu z kamery bychom tradičním způsobem potřebovali vyřešit nespočet dílčích problémů (zpracování obrazu, rozpoznání, lokalizace a klasifikace objektů, silnice a jízdních pruhů, značek nebo ostatních aut atd., předpověď trajektorie, výpočty pro ovládání směru jízdy a nastavování rychlosti, korekce chyby, předvídaní události v provozu, pohybu a intencí ostatních aktérů atd.), jedna dobře nadesigovaná a natrénovaná architektura ANN to může teoreticky zvládnout celé zároveň<sup>15</sup> – máme v tom případě mluvit o ANN jako o množství algoritmů?

---

<sup>14</sup> viz např. AlphaGo (šachová verze)

<sup>15</sup> Jediný problém spočívá v tom, že pak nelze jednoznačně určit, kde jednotlivé „algoritmy“ začínají a končí – například co je ještě otázka zpracovávání obrazu, tedy „percepce“, a co už řízení, tedy akce – případně jestli se to neprolíná a vozidlo není ovládáno na základě nějaké těžko popsatelné obrazové intuice. To je jeden z důvodů, proč se prozatím tato metoda (jedna ANN na všechno) v této oblasti nepoužívá.

To by jako nejbližší analogii pro ANN znovu asociovalo model počítače. V podstatných věcech se sice ANN liší, neobsahuje například systematicky uspořádané komponenty (procesor, paměť atd.), přesto je možné to tak interpretovat – pomocí modelu tzv. paralelního distribuovaného zpracování.<sup>16</sup> Pokud to vezmeme z hlediska mikrosvěta, tedy srovnáním struktury a zapojení ANN třeba s zapojením tranzistorů v počítači, narazíme rovněž na zajímavé podobnosti, ale i zásadní odlišnosti. Vhodným uspořádáním neuronů by bylo teoreticky možné funkci elektrického obvodu (například logických hradel) simulovat<sup>17</sup> a naopak, možná že velkou část fungování NN obecně by modelem netriviálního obvodu bylo možné objasnit (a nutno podotknout, že pokud ANN běží jako program na počítači, vlastně se její procesy do podoby logických operací na úrovni tranzistorů přenášejí).

„Obvod“ ANN nicméně funguje na odlišných principech. Jednak není nijak systematický – každý neuron se spojuje se všemi ostatními neurony v předchozí a následující řadě (a v některých modelech i zpětně nebo napříč) a přenos signálu po synapsích funguje probabilisticky (signál není binární, ale v množině reálných čísel). Navíc až v procesu učení konkrétní zapojení „vykrytalizuje“ – některé synapse dostanou větší váhu a jiné menší, takže přenáší signál s různou intenzitou. Také tento „obvod“ obsahuje vysoké množství redundance – to je dáno synapsami či neurony, které přenáší zanedbatelný signál (respektive neovlivňují výsledek). V logice „obvodu“ bychom to chápali nejspíš jako známku neefektivity, naopak se to však ukazuje jako vlastnost, díky které model v lecčemž překračuje limity tradičních metod.

## Struktura ANN a BNN

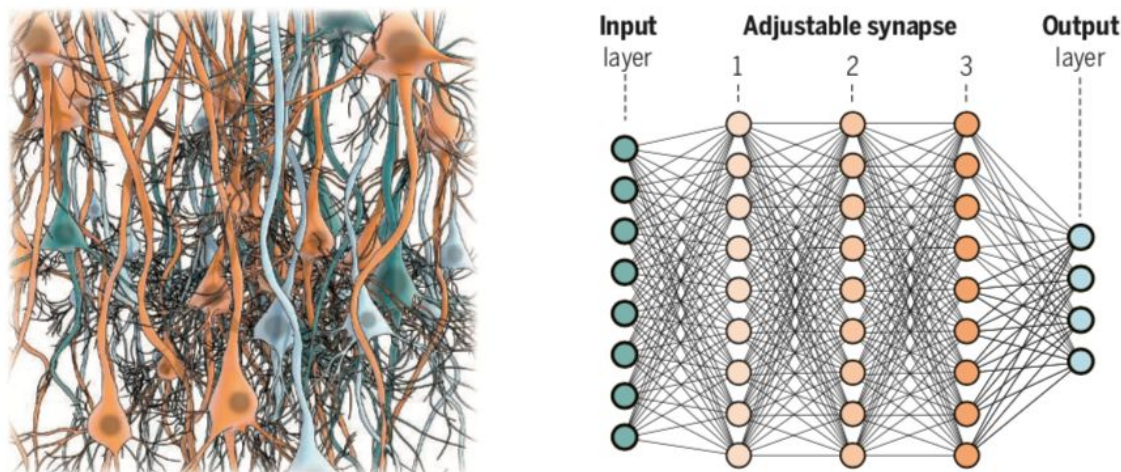
Čistě topologicky je ANN pořád nejbližší struktuře BNN, což pravděpodobně není příliš překvapivé, nicméně u ANN, narozdíl od BNN, se intuitivněji uvažuje i o souvislostech z hlediska výpočetní techniky a teorie informace – proto nehodlám opomíjet ani jednu ze zmiňovaných modalit.

---

<sup>16</sup> [doplnit citaci]

<sup>17</sup> Z analýzy výpočetní síly různých architektur ANN celkově vyplývá, že mohou dosahovat téměř jakékoliv úrovně, až po kapacitu tzv. univerzálního Turingova stroje, nebo dokonce po superturingovskou výpočetní sílu, viz: Balcázar, José (Jul 1997). "[Computational Power of Neural Networks: A Kolmogorov Complexity Characterization](#)". *Information Theory, IEEE Transactions on.* **43** (4): 1175–1183.

## Ilustrace – Topologie ANN a BNN



// Obr.1 srovnání ANN (vpravo) a BNN (vlevo).<sup>18</sup> Hlavním rozdílem ve struktuře mezi ANN a BNN je topologická složitost, zatímco ANN pracuje s limitovaným počtem velmi zjednodušených umělých neuronů, jež jsou uspořádány v homogenních vrstvách, reálná neuronová síť se vyznačuje bohatým propojením, včetně lokální a dálkové, laterální konektivity nebo zpětných [top-down] spojení z vyšších do nižších úrovní hierarchie. [viz Ullman] Kromě komplexnější morfologie do funkce biologických neuronových sítí zasahují také různé fyziologické efekty a neurochemické procesy (např. neurotransmitery). Nicméně přes všechna zjednodušení a neshody (Ullman pozn. 5) ukazuje srovnání některých modelů (pomocí metod empirické psychologie nebo f-MRI skenování), že existují i překvapivé podobnosti (Ullman pozn. 4), obzvláště na nižších úrovních hierarchie sítě.<sup>19</sup>

—

Vidíme, že na základě struktury tak ANN a BNN lze z velké části srovnávat. To už neplatí například o procesu učení nicméně nezávisle na tom jak se synapse NN nastaví (zda-li trénováním metodou Backpropagation, evolučními algoritmy, nebo jakkoliv jinak), na strukturálních podobnostech to nic de facto nemění. Nicméně není vhodné srovnávat to obecně – struktury v rámci ANN se svým uspořádáním liší, liší se jednak architekturou určenou pro zpracovávání konkrétního typu dat (obraz, zvuk, text atd.) a jednak volbou hyperparametrů, jež se přizpůsobují konkrétní úloze. V některých případech se používají také soubory více provázaných ANN, ať na bázi

<sup>18</sup> viz Ullman 2019 science neuroscience  
<http://science.sciencemag.org/content/363/6428/692>

<sup>19</sup> Což je logické – nižší úrovně hierarchie sítě zachycují obecnější ...

paralelní funkce, sériového navázání, kooperace, nebo dokonce soupeření.<sup>20</sup> Lze očekávat, že se budou obdobně lišit i struktury BNN.

Různé druhy organismů mají vrozené specifické mozkové struktury (tedy struktury BNN) dedikované pro konkrétní účely – například některé ryby nebo ptáci dokáží pomocí nich řešit složité navigační úlohy, které by jinak bylo extrémně náročné se bez genetických vloh naučit.<sup>21</sup> Přičemž tyto mozkové struktury, které usnadňují osvojení si komplexních intelektuálních schopností, se pravděpodobně formují evolučně a podléhají také potřebám společenství. BNN zde funguje v podstatě jako paměťové médium – z generace na generaci přenáší určitou zkušenost (například zkušenost „navigace“) ve formě samotné struktury. Do stejné kategorie bychom u člověka mohli zařadit i zkušenost estetickou (viz kapitola Percepce a imaginace). Pravděpodobně jakákoliv struktura NN, včetně ANN, v sobě inherentně skrývá určitou informaci, „zkušenost“ či „paměť“, jež se potom může projevat při zpracovávání dat. V tomto kontextu se nabízí nadnést i otázku vrozených jazykových dispozic, jak ji postuloval Chomsky. Předtím je ale záhodno zabývat se kompatibilitou těchto odlišných struktur vůbec (kompatibilitou ANN, BNN, jazyka, nebo dalších zmiňovaných interpretací).

## Percepce a imaginace

Obávám se, že rozbor statistických vlastností nebo algebraických vztahů na pozadí ANN by nevedl k žádnému přesvědčivému objasnění principu jejich fungování – chceme-li ANN interpretovat, nikoliv pouze předložit fakta, jak to činí technická literatura, nutné je zvolit jiný jazyk. V této kapitole se tak budu primárně zabývat otázkami: Jakou má ANN perspektivu? Jakým způsobem nahlíží na svět – na čem zakládá svůj názor? Co charakterizuje její úhel pohledu?

---

<sup>20</sup> S touto metodou se setkáme například v architekturách GAN (generative adversarial networks). V architektuře GAN se neuronové sítě trénují vzájemně tím, že spolu „soupeří“ – jedna se snaží překonat výsledek té druhé a naopak. Toto soupeření obvykle vede k překvapivým výsledkům, především co se týče vytváření originálních výsledků a nových strategií. <https://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>

<sup>21</sup> viz Ullman 2019

Vizuální metafory míním doslova i obrazně – zabývat se dále budu primárně CNN, tedy modely uzpůsobenými na zpracovávání obrazu. Důvod je ryze praktický – aby bylo co vidět – obzvlášť, když se naskýtá možnost vnitřní mechanismy ANN přímo sledovat. Metaforami také záměrně juxtaponuji otázku vidění a obecného zpracovávání dat – jak ještě uvidíme, vizuální architektura CNN se velmi podobá ANN řešícím intelektuálně náročné úlohy (např. AlphaGo Zero) – je algoritmické „myšlení“ založené na vidění, podobně jako když Platon mluvil o poznání ve světelných metaforách a v metaforách vidění?

Věřím, že teprve skrze rozbor vnímání, jakým je také vizuální percepce CNN (ANN určených pro zpracovávání obrazu), kterou se budu v této kapitole primárně zabývat, je možné hlouběji porozumět reálné „zkušenosti“ těchto modelů, tedy vnitřním mechanismům, které se v nich odehrávají, ale zatím nejsou nijak uspokojivě interpretovány. Konfrontace s touto neznámou a navýsost bizarní estetikou navíc může vést k podstatným otázkám o zkušenosti lidské. Nejprve představím metodu a umělecký směr inceptionismu, ze kterého pochází dále rozebírané ilustrace a následně se budu zajímat již čistě jejich estetikou.

## Inceptionismus

S technikou inceptionismu<sup>22</sup> nebo také programem DeepDream přišel softwarový inženýr Googlu Alexander Mordvintsev. Využívá efektu algoritmické vizuální pareidolie (dovytváření nezřetelných podnětů ve smysluplné obrazy za pomoci fantazie) ke zvýraznění obrazových vzorů a produkci snové či halucinogenní estetiky.<sup>23</sup> V psychologii se jedná o známý jev – jako příklad se uvádí rozpoznávání tvarů v oblacích – člověk v nich dokáže rozpoznat obrysy tváří, zvířat nebo jiných objektů, i když jde o velmi vzdálenou podobnost.<sup>24</sup> Nicméně v případě výpočetního modelu – u kterého se nic jako fantazie nepředpokládá – jde o poměrně unikátní fenomén. Aby bylo jasnější, jak může k podobnému efektu docházet, je nutné

---

<sup>22</sup> název podle filmu Inception a stejnojmenné architektury ANN

<sup>23</sup> Mordvintsev, Alexander; Olah, Christopher; Tyka, Mike (2015). "[DeepDream - a code example for visualizing Neural Networks](#)". Google Research. Archived from [the original](#) on 2015-07-08.

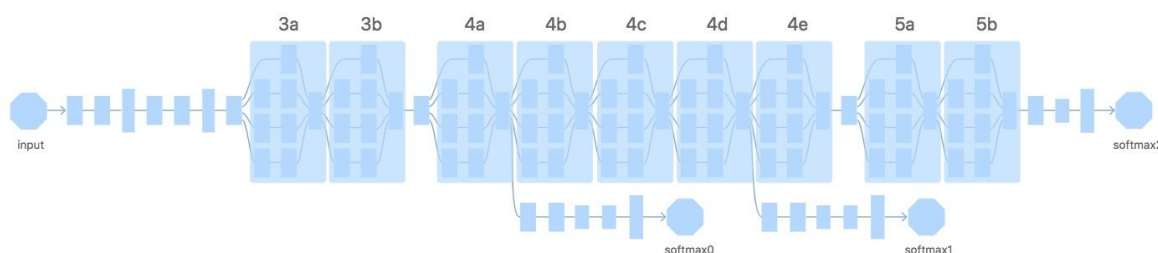
<sup>23</sup> \_a\_ Mordvintsev, Alexander; Olah, Christopher; Tyka, Mike (2015). "[Inceptionism: Going Deeper into Neural Networks](#)". Google Research. Archived from [the original](#) on 2015-07-03.

<sup>24</sup> [VONDRÁČEK, Vladimír](#). *Fantastické a magické z hlediska psychiatrie*. Bratislava: Columbus, 1993. S. 54.



nejprve porozumět principu CNN a jeho podobnosti s lidskou vizuální kůrou. Některé z aspektů nejprve představím na původní CNN – GoogLeNet (nazývané také Inception).

## GoogLeNet



//Distill – GoogLeNet//

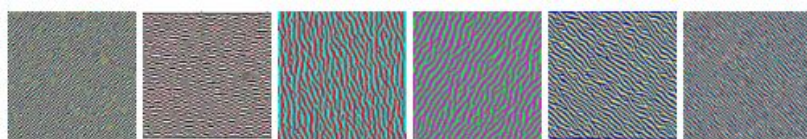
Na stránkách projektu Distill je interaktivní vizualizace charakteristik sítě GoogLeNet.

<sup>25</sup> Ta má následující architekturu (obr.XY). Oddíly 3a-5b jsou vrstvy neuronů, které na základě různých naučených znaků [features] klasifikují vstupní fotografie. Děje se tak pomocí procesu konvoluce<sup>26</sup> – fotografie se pokaždé přefiltruje pomocí konvoluční masky (tu můžeme interpretovat jako menší fotografii) a výsledná odezva indikuje, jaké části fotografie odpovídají „hledanému“ znaku na konvoluční masce. Například pokud na fotografii hledáme hrany, zvolíme konvoluční masku s ostrým přechodem a výsledkem budou detekované hrany na původní fotografii – analogicky lze detekovat i složitější objekty. Aby byla CNN schopná efektivně rozpoznávat obrázky, musí nalézt odpovídající kritéria [features] – a k tomu konvoluční masky slouží: CNN pomocí nich v jednotlivých vrstvách neuronů rozřazuje obrázky, od nejabstraktnějších kritérií – hrany a linie, vzory a textury, základní tvary – až po ty nejkonkrétnější – části objektů, celky nebo složitější kompozice. Takto funguje princip hierarchické reprezentace.

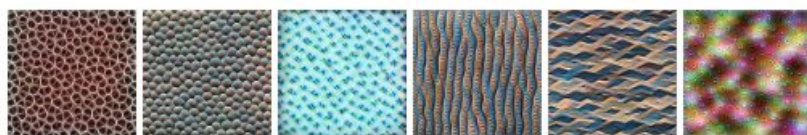
<sup>25</sup> <https://distill.pub/2017/feature-visualization/appendix/> (Přístup duben 2019)

<sup>26</sup> Mimořádně kvalitní a názorné vysvětlení v kontextu CNN nabízí Alexander Colah: <http://colah.github.io/posts/2014-07-Understanding-Convolutions/> (Přístup duben 2019)

## Hierarchická reprezentace



Edges (layer conv2d0)



Textures (layer mixed3a)



Patterns (layer mixed4a)



Parts (layers mixed4b & mixed4c)



Objects (layers mixed4d & mixed4e)

//Distill – Vizualizace kritérií [features], na základě kterých se GoogLeNet rozhoduje. Od shora: Hrany (vrstva conv2d0), Textury (vrstva mixed3a), Vzory (vrstva mixed4a), Části (vrstvy mixed4b & mixed4c) a Objekty (vrstvy mixed4d & mixed4e).

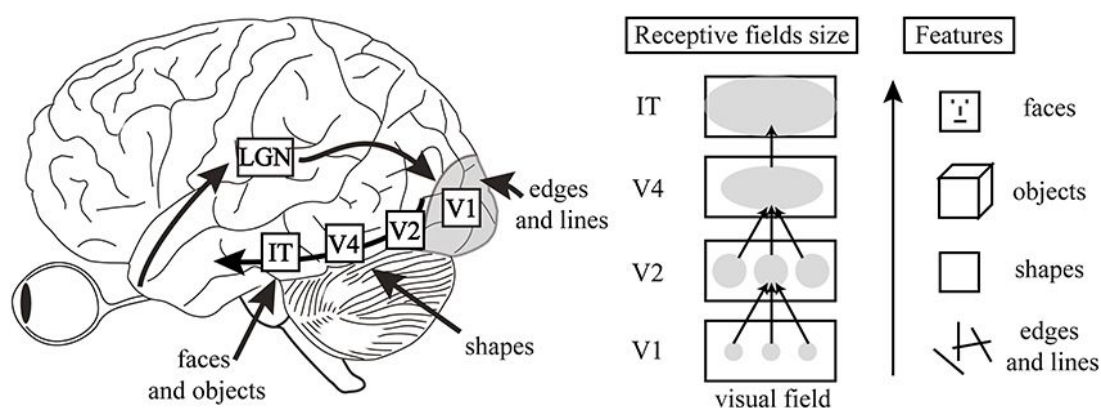
První vrstvy rozlišují základní tvary a textury – křivky a linie, vlny, rohy, zakulacení atd. – a další vrstvy vnímají čím dál složitější tvary, abstraktnější souvislosti a konkrétnější objekty. Neurony v druhé vrstvě kupříkladu rozlišují něco jako fleky a puntíky, linie, které připomínají rostlinné tvary nebo vlnky. Ve posledních vrstvách už vidíme kritéria jako psí uši nebo architektonické oblouky. Podobných kritérií je ve skutečnosti mnohem více, ale čím jich síť potřebuje méně, tím funguje efektivněji – vytváří tedy taková, která jsou na sebe pokud možno ortogonální a co nejvíce se odlišují. //

Princip hierarchické reprezentace vypovídá něco o distribuci významu v ANN – nefunguje to tak, že každý neuron reprezentuje jednu konkrétní informaci, například, že bychom jako ve slovníku v jednom neuronu našli onu ideu psa, která se

aktivuje, když síť na obrázku psa detekuje. Naopak, obrázek psa aktivuje celou řadu neuronů reprezentujících texturu srsti, barvu daného plemene, oči, uši, končetiny nebo třeba travnatý povrch a na výstupu poté zjistíme, že síť vidí například z 73% dalmatina, z 15% kočku, z 10% pole a z 2% rohožku – významy tak nejsou nijak definitivní a kategorie nejsou neprostupné nebo nějak jasně vymezené, zachycují v podstatě sémantickou syntézu dílčích aspektů, které potom dohromady reprezentují konkrétní věc. Každý význam je zároveň určen i strukturou všech ostatních významů a vizuální vjem je probabilistickou kombinací všech asociací, což si lze představit jako abstraktní vektor.<sup>27</sup> Podrobněji se tomu budu věnovat ještě v kapitole Emergence a distribuovaný význam na příkladu modelu ANN pracujícího s jazykem.

### Lidská vizuální kůra

Důležitým poznatkem ohledně principu hierarchické reprezentace je, že na jeho základě funguje i vizuální kůra savců. To myslím samozřejmě s obrácenou kauzalitou – modely jako CNN jí byly inspirovány.<sup>28</sup> Struktura vizuální kůry se tak struktuře CNN nápadně podobá.



Převzato ze studie <http://www.jneurosci.org/content/jneuro/35/27/10005.full.pdf>

Kromě toho, že CNN v rozpoznávání obrazu dosahuje srovnatelných výsledků s člověkem, dosahuje jich na základě stejné vnitřní struktury a podobných mechanismů – některé studie porovnávají vrstvy neuronů v CNN s konkrétními

<sup>27</sup> Nicméně i to lze vizualizovat – viz Distill <https://distill.pub/2018/building-blocks/>

<sup>28</sup> CNN má kořeny v modelu Neurocognitron <https://www.rctn.org/bruno/public/papers/Fukushima1980.pdf> který explicitně usiloval o napodobení vizuální kůry.

oblastmi vizuální kůry (V1, V2, V4 nebo IT), nebo dokonce na základě CNN předvídají mozkovou aktivitu skenovanou f-MRI.<sup>29</sup> Největší podobnosti byly zjištěny v předních vrstvách CNN a vizuální kůry, tedy tam, kde jde o rozpoznávání hran a linií, textur nebo vzorů.

### Psi a pagody | Ilustrace inceptionismu

Na následující ilustraci jsou záběry s dvou videí Johana Nordberga „Inside an artificial brain“ (vlevo) a „Inside another artificial brain“ (vpravo), jež nabízejí poměrně reprezentativní vhled do vnitřních procesů CNN pomocí softwaru DeepDream.

Norberg použil CNN natrénovanou na rozpoznávání velkého množství obrazů, ale namísto normálního obrazu dal na vstup náhodný šum a síť nechal, aby zvýrazňovala rysy toho, co rozpoznává – jedná se tak o jakýsi spontánní proud představ. Model z prvního zmiňovaného videa je natrénován na všeobecné databázi ImageNet a ten druhý na databázi obrázků míst MIT Places. Proces na videu probíhá stejně jako proces rozpoznávání vrstvy neuronů po vrstvě – zatímco první vrstvy jsou citlivé na body, linie a základní vzory, v pozdějších vrstvách už jde o komplexní koláže celých objektů (viz podkapitola Hierarchická reprezentace). Přiložené ilustrace jsou z pozdějších vrstev po druhé minutě videa, kdy se začínají obrazy proměňovat v poměrně překvapivé halucinace.



Dvě videa Johana Nordberga<sup>30</sup> „Inside an artificial brain“ (vlevo) <https://vimeo.com/132700334> a „Inside another artificial brain“ (vpravo) <https://vimeo.com/153297242> (2015) nabízejí poměrně

<sup>29</sup> Viz například <http://www.jneurosci.org/content/jneuro/35/27/10005.full.pdf> nebo <https://hal.inria.fr/hal-01389809/document> (Přístup duben 2019)

<sup>30</sup> <https://johan-nordberg.com/> (Přístup duben 2019)

reprezentativní vhlad do vnitřních procesů CNN pomocí softwaru DeepDream

<https://ai.googleblog.com/2015/07/deepdream-code-example-for-visualizing.html>

–

## Psi a pagody \_ Dilema

Otázka zní takto: Čím to je, že skryté vrstvy neuronů v CNN zobrazují v takové míře psi a pagody? Jistě, lze uvést prozaickou hypotézu – v datasetu byl velký počet obrázků na toto téma (databáze zjevně čerpají z internetu, kde zkrátka převládají fotografie domácích mazlíčků a turistických destinací)... Ale proč to vypadá tak divně? Proč namísto ideje psa, pomyslného abstraktního předobrazu toho, co psa definuje, vidíme spíše divokou halucinaci monstrózních bytostí s desítkama očí a končetinama na hlavě? Nonsensuální koláž, která se navíc v původním datasetu rozhodně vyskytovat nemohla! Co to má znamenat, jestliže se jedná o algoritmus, který by měl fungovat systematicky a rozhodovat na základě objektivních kritérií?

–

## Fenomenologie a psychoanalýza CNN

Na základě úzké souvislosti mezi CNN s lidskou vizuální kůrou je třeba CNN chápat jinak – nikoliv jako arbitrární výpočetní model, ale jako abstraktní subjekt jistého vizuálního podvědomí. Toto podvědomí se pravděpodobně podobá tomu lidskému, nebo funguje na základě srovnatelných principů; jelikož se ovšem nejedná o věrohodnou „kopii“ BNN, zato o zajímavou „simulaci“, není možné ho s vizuální kůrou konkrétního jednotlivce ztotožnit, ale je nutné přisoudit každé architektuře CNN unikátní „zkušenost“. Ačkoliv jde o ryze vizuální aparát zproštěný jakýchkoliv fenomenologických kategorií – tělesnosti, zkušenosti času, starosti, vrženosti atd., domnívám se, že k plnohodnotné filosofické diskusi postačí kategorie (byť nevědomého) vnímání. Ba dokonce, vytrženost vizuálního vnímání z kontextu dalších kategorií může být v případě analýzy vizuálních halucinací, snů a skrytých představ v konečném důsledku možná i podnětější – můžeme si být například jisti, že sny CNN nemají žádnou spojitost se sexualitou, agresí nebo jinými aspekty, jež pojmenovává Freud, nýbrž vyplývají čistě ze zkušenosti vnímání.

Když Merleau-Ponty přednáší o „primátu vnímání“,<sup>31</sup> představuje to pro něj způsob jak přistupovat k filosofickým otázkám obecně. V konečném důsledku lze totiž na většinu problémů – pojmy jako čas, prostor, řeč, rozum atd. – nahlížet jako na problémy vnímání. Vnímání jako předpoklad konání – vjem, který je prvotní, časově předchází řeči nebo myšlení a zároveň představuje ryzí zkušenost „věcí samých“ jak se zobrazují v našem vědomí. Vnímání jako empirie – epistemologický fundament, předpoklad jakéhokoliv poznání. A nakonec vnímání jako analogie pro filosofickou metodu. Ve všech těchto významech lze primát vnímání chápat. V případě CNN mi ovšem přijde trefnější Merleau-Pontyho definice vnímání v kontextu umělecké tvorby:

*Malířova vize už není pohledem upřeným navenek, jen „fyzikálně optickým“\*\*\*\*\* vztahem ke světu. Svět už není před malířem jako představa, nýbrž spíše malíř sám se rodí ve věcech jako koncentrace viditelná, jež si uvědomilo sebe sama, takže obraz se nakonec nevztahuje k ničemu mezi empirickými věcmi, pokud jde zprvu o obraz „autofigurativní“. Nezobrazuje nějakou věc, ale je „zobrazením ničeho“, roztrhává „kůži věcí“, aby ukázal jak se věci stávají věcmi a svět světem. [...] Umění je vpravdě „neartikulovaným výkřikem“, jaký vydával Hermés Trismégistes, který „se zdál hlasem světla“.*<sup>32</sup>

Vnímání CNN a inceptionismus je nutné také chápat jako „zobrazení ničeho“ či „neartikulovaný výkřik“ – a to je domnívám se hlavní důvod, proč dává smysl se tímto tématem zabývat.

Psi a pagody \_ Interpretace

Inceptionistickou estetiku je nutné analyzovat na více úrovních abstrakce – postupovat budu skrze analýzu jednotlivých aspektů od úrovně vzorů a textur přes jednotlivé části až po výslednou kompozici. Není cílem pouze zodpovědět dilema

---

<sup>31</sup> Merleau-Ponty, M., Primát vnímání a jeho filosofické důsledky, Togga, Praha 2011

<sup>32</sup> Merleau-Ponty, M., Oko a duch a jiné eseje, Obelisk, Praha 1971, str. 26

psů a pagod, ale nějakým způsobem charakterizovat obecné rysy inceptionismu, které zároveň vypovídají o vlastnostech ANN obecně a v neposlední řadě se dobrat zajímavých podobností či odlišností s člověkem.

#### Dynamika a ornamentalita

Imaginace CNN na nižších úrovních abstrakce, kde zvýrazňuje vzory, křivky a linie, zlomy a zaoblení, umožní nejlépe demonstrovat obraz jako je Nedělní odpoledne na Le Grande Jatte od Geogrese Seurata. Všimněme si, že pointilistickou plochost a bezpříznakovost CNN transformuje v plastickou ornamentalitu – ať už spojením teček v pomyslné tahy štětcem (obr.1), pomocí systematického šrafování vzdáleně připomínajícího Escherovy deformované geometrie (obr.2), nebo dokonce skrze zákruty a spirály (obr.4).



Vizualizace vytvořené softwarovým inženýrem Googlu *Matthew McNaughtonem*.<sup>33</sup> *Originál je Georges Seurat Nedělní odpoledne na Le grande Jatte, zpracování pak je zřejmě z jedné z předních vrstev stejné úrovně abstrakce, takže se liší pouze dílčí způsoby, jakými CNN obrazy zpracovává. Označuji je následujícím způsobem. první řada zprava: 1, 2 a druhá řada zprava: 3, 4*

Ve všech případech jde o mimořádně dynamické zobrazení – spirály evokují turbulentní proudění vangoghovského impresionismu, linie na sebe plynule navazují a přechází jedna do druhé podle zásad ornamentální estetiky – v pozadí se ztrácí rozdíl mezi siluetou člověka, kmenem stromu nebo konturami země, objekty se rozplývají v abstraktní tvary. Celek tím vyvolává dojem pohybu a neustálé proměnlivosti, čehosi živelného, organického (jako textura na obr. 3).

Dynamické vzezření bychom mohli pokládat vedlejší efekt zmiňované ornamentality, nicméně nabízí se i zajímavější interpretace. Flusser popisuje jako jeden z předních aspektů vnímání fotografie jisté magično.<sup>34</sup> To je podle něj tvořeno zploštěním reality ze čtyř dimenzí do dvou, tedy odebráním kauzálních vztahů, které si poté člověk má tendenci doplňovat.<sup>35</sup> Jinak řečeno, fotografie zastavuje čas, odebírá realitě pohyb a plynutí a člověk má potom tendenci si to kompenzovat, statické vnímat dynamicky – vidět ve fotografii akci a číst jí jako příběh s vlastní kauzalitou. Nutno podotknout, že tohoto efektu využívá v hojné míře komix – juxtapozicí jednotlivých „pohledů“ vytváří filmový efekt, pokud se však pohled zastaví, akce neustává, ale pokračuje i v jednom statickém rámečku. Ačkoliv víme, že CNN si žádnou kauzalitu nedoplňuje, nemá zkušenost času ani prostoru a vnímá pouze dvě dimenze, přesto tu efekt podobný

---

<sup>33</sup><https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html> (Přístup duben 2019)

<sup>34</sup> Momentálně sice není řeč o fotografii, nicméně Seuratův pointilismus se jí velmi podobá – především „pixelací“ ploch do teček jednotlivých barev, které potom člověk vnímá jako jednolitě, přestože na sebe nijak nenavazují.

<sup>35</sup> *Zatímco pohled ohledávající obrazovou plochu zachycuje jeden prvek po druhém, vytváří mezi nimi časové vztahy. Může se vracet k již viděnému prvku obrazu, a z „předtím“ se stává „poté“: Čas rekonstruovaný scanningem [ohledáváním povrchu obrazu] je časem věčného návratu téhož. [...] Tento časoprostor, který je obrazům vlastní, není ničím jiným než světem magie, v němž se všechno opakuje a v němž má vše svůj podíl na jakési závažné souvislosti celku. Takový svět se strukturálně odlišuje od historické lineárnosti, v níž se neopakuje nic a v níž má vše své příčiny a následky. Například v historickém světě je východ slunce příčinou kokrhání kohoutů, v magickém světě východ slunce znamená kokrhání a kokrhání východ slunce. Význam obrazů je magický. (Flusser, *Technický obraz*, 11-12)*



Flusserovu magičnu spatřujeme. Možná je již součástí samotného pohledu – přestože zrak mezi ostatními smysly tradičně evokuje odstup, zastavení a možná onu zkušenost fotografie, tohle by znamenalo, že vizuální vnímání zaujímá naopak aktivní roli a pro samotné rozpoznání základních vzorů a textur je nutné vnímat pohyb, respektive vnímat v pohybu. Zajímavé to je ve chvíli, kdy si uvědomíme, že nejde o výplod přebujelé fantazie, nýbrž o neekonomičtější způsob klasifikace obrazu, ke kterému CNN došla – formu určité abstrakce a schematizace.

### Fraktální geometrie

Aspekt ornamentalit je ještě intenzivněji vidět v předních vrstvách videí „Inside an artificial brain“, ze kterých pochází přiložené ilustrace psů a pagod. Už ovšem nejde o zvýraznění předlohy jako v případě Seuratova obrazu, ale o čistou imaginaci. Také se jedná o trochu pozdější vrstvy, jelikož vzory a ornamente nabývají na složitosti a postupně splývají v celky. Všimněme si, jak komplikovaný ornament na obrázku vlevo volně přechází z abstraktních tvarů, zákrutů a spirál v kontury psiho obličejce. Na obrázku vpravo splývá geometrie kumulujících se architektonických oblouků s zákrutem voluty nebo obloukem do tvaru ochozu kolonády. Celek je nicméně navýsost ambivalentní. Kromě prolínání forem není možné rozlišit mezi celkem a částí, mezi konkrétním a abstraktním.



//Psi a pagody, záběry z dřívějších vrstev//

Záběry z tohoto „meziprostoru“ rovněž ukazují, že mezi vnímáním vzorů a textur nebo celých objektů není u CNN žádný ostrý přechod – už v ornamentu se rýsují

kontury celku a zdá se, že celky – jako na ukázkách psů a pagod z počátku kapitoly – dohromady můžou představovat pouze komplikovanější ornament. Minimálně kompozice tomu naznačuje – odvíjí se od jednotlivých vizuálně intenzivních znaků (oči, špičky pagody atd.) a zbytek je vyplněn směsicí textur a abstraktních kontur, ze kterých můžou postupně vykristalizovat další objekty. Jinak kompozici charakterizuje absence jakékoliv hierarchie a kolážovitost umožňující číst obraz ze všech stran. Kompozice vychází z ornamentu a rekurzivně se k němu navrácí – jinak řečeno, složitých tvarů se dosahuje opakováním jednoduchých pravidel. To vysvětluje komplexnost představ CNN – jsou emergentním jevem – ucelený vjem vytváří systematická, leč spontánní interakce jednotlivých vzorců, které pak podobně jako písmena v abecedě vytvářejí neomezené množství složitějších kombinací určujících vlastnosti celku. Fenoménem emergence se budu více zabývat v kapitole Emergence a distribuovaný význam.

Když nad tím zauvažujeme, fraktální geometrie, která se v těchto představách formuje, je co se týče vizuálního vnímání mimořádně efektivní strategií. Jedná se o předpoklad, že se ty samé vzory neustále opakují nezávisle na měřítku. To je aspekt, který definuje estetiku všech přírodních úkazů, které tak podněcují lidskou fantazii – oblak, lesů, skal, mořské hladiny atd. Fraktální geometrie, to je pro člověka údajně jedno z univerzálních estetických kritérií.<sup>36</sup> Co je na fraktálech tak působivého?

Estetiku fraktálů nejintenzivněji pociťujeme v gotice a romantismu. Poněvadž fraktály navozují pocit vidiny nekonečna, rozlohy bez hranic – jsou to nedozírné dálky, bezedné propasti a nepřístupné skály na obrazech Caspara Davida Friedricha. Mlhavá údolí, jeskyně obrostlé mechem, poetika hor a estetika ruin. Prožitek vznešena – pohledy snoubící hrůzu a úzkost – zkušenost exotiky, napětí a překvapení.

---

<sup>36</sup> Spehar, B., Clifford, C. W. G., Newell, B. R., & Taylor, R. P. (2003). *Universal aesthetic of fractals. Computers & Graphics, 27(5), 813–820.*doi:10.1016/s0097-8493(03)00154-7

Pocity, které tato estetika navozuje – dojem nepopsatelného a nevyslovitelného, potažmo transcendentálního – jsou zážitkem hranic lidského vnímání a poznání. Dálky, do kterých nelze dohlédnout, mlhavé hřebeny hor, které nelze rozeznat, jeskyně a zákoutí – nedotčené krajiny navždy skrývající tajemství. V tom přichází na řadu imaginace. Fraktály nabízí nekonečné možnosti interpretace – a to člověka pohlcuje, protože vnímá věci na hranici vnímatelného – snaží se rozeznat to, co rozeznat nelze a obrací se tak do svého nitra, které tyto vjemy utváří – vznešeno přítomné v romantické zkušenosti utváří konfrontace se sebou samým, kontakt s proudem snů, podvědomých představ a tužeb, který probouzí fraktální estetika.

–

Fraktály ve vnímání CNN hrají výsadní roli z několika důvodů. Jedním z nich je zmiňovaný předpoklad nekonečných možností interpretace – v každém detailu se může ukrývat nový kosmos – to probouzí proud asociací, potenciálních tvarů, jež mohou předznamenávat přítomnost skrytého, očekávání nového, neznámého a překvapujícího – jakákoliv část potenciálně obsahuje celek. Na přiložených ilustracích jakoby střecha pagody byla složená z nekonečna dalších možných pagod nebo se tvář psa skládala z nekonečna dalších možných tváří. Co se týče architektonických tvarů, působí to přirozeně – těžko bychom také hledali fraktálnější architekturu než jakou nabízejí pagody – v druhém případě to však působí děsivě. Roztříštěnost organických částí, schizofrenní všudypřítomnost očí, tělesné výběžky a otvory, rozplízlé tvary konstituující bezforemná monstra.

Paranoia, naivita, sen a grotesknost

Pokud se přesuneme do vyšších vrstev abstrakce – z úrovně textur, hran a linií, do úrovně větších celků a objektů – na základě předchozí kapitoly můžeme položit otázku, jestli je dynamika, ornamentální esteta nebo fraktální esteta pro vidění CNN příznaková i na těchto vyšších úrovních, případně kam se posouvá.



"Admiral Dog!"



"The Pig-Snail"



"The Camel-Bird"



"The Dog-Fish"

*Po přiblížení můžeme identifikovat různé opakující se útvary, jako například 'Pes admirál', 'Praso-šnek', 'Velbloudo-pták' nebo 'Rybo-pes'. V trénovacím datasetu této neuronové sítě se pravděpodobně nacházelo hodně fotek zvířat.*<sup>37</sup>

Na ilustraci výše vidíme několik ukázek útvarů, které si CNN představuje v oblacích. Jsou to bájně bytosti ze středověkých bestiářů, nebo groteskní stvoření na motivy

---

<sup>37</sup> <https://ai.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html> (Přístup 2019)

karnivalové estetiky? Stejně tak bychom mohli uvažovat o surrealismu nebo o snových vizích z triptychu Hieronyma Bosche.

V těchto případech ornamentální a dynamika přesahuje v grotesku – ryba plynule přechází v psa, šnek do tvaru prasete atd. – to doplňuje výrazný nepoměr tělesných částí (zejména končetin a torsa) a samozřejmě absence jakékoliv hierarchie nebo struktury – bytosti mají všelijak rozmístěné oči a ústa, někdy není vůbec jasné, co daný shluk tvarů připomíná nebo reprezentuje. Groteskní pocit navozuje také analogie na základě tvarové podobnosti (například psí čumák, jež tvoří zároveň hlavu ryby), jde o rafinované vizuální metafory, nebo pouze náhodné kombinace častých tvarů zakódovaných v datasetu, na kterém se ANN trénuje?

Jakákoliv interpretace vizuálního vjemu musí nutně něco zvýrazňovat a dotvářet, zatímco ostatní opomíjet. Stejně to funguje v libovolném případě zpracovávání dat (viz No free lunch theorem). V obrazech se to nicméně projevuje nejnázorněji. Jedná se především o dva mechanismy, které se v problému regrese nazývají underfitting a overfitting (viz Appendix, Volba hyperparametrů) – underfitting znamená podcenění složitosti dat a přílišnou aproximaci, overfitting naopak přecenění složitosti dat a nedostatečnou generalizaci – oba extrémů vedou k nižší výsledné přesnosti. Neformálně bych underfitting nazval jako naivitu a overfitting jako paranoii. Na inceptionistických ukázkách je to myslím velmi názorné – jelikož obrazy charakterizuje tak velká datová komplexnost (není divu, že automatizované rozpoznávání obrazu byla dlouho spíše záležitost sci-fi imaginace než realistických vyhlídek v oboru počítačového vidění), ANN využívá obou principů – v nižších vrstvách abstrakce (viz Seurat) dochází primárně k seskupování roztříštěných částí do ucelených tvarů, linií a vzorů, zatímco ve vyšších vrstvách abstrakce se už tvary a části různých objektů už pouze nekontrolovatelně kumulují. Bez určité dávky naivity není možné vnímat na obraze podobné – uvědomovat si barevné plochy, kontury a siluety – a odlišné a cizí interpretovat jako známé, jež je součástí celku; bez paranoie naopak přehlídíme detaily, skryté vzory a významy, opomíjíme jednotlivost a odlišnost, podceňujeme rozmanitost, nekonzistenci a nepředvídatelnost reality. Obvykle s nedostatkem podnětů inklinujeme k paranoie (například v noci či ve snu) a s jejich přemírou naopak k naivitě.

Působí to, jakoby paranoia byla pro vnímání něčím fundamentálním – zjevně protože funguje konstruktivně, dotváří obraz, doplňuje do něj představy a asociace, potenciální tvary a jejich části, dodává vjemu dynamiku a nachází nečekané souvislosti. Myslím, že se to podobá mýtické zkušenosti nebo Flusserově magičnu. Především jde ale o předpoklad jakéhokoliv vnímání, jelikož každý vjem je interpretován teprve představami a asociacemi, které ho formují.

#### Exkurz \_ Groteskní omyly aneb Efekt motýlích křídel

Ačkoliv je paranoidní strategie CNN mimořádně efektivní co se týče rozpoznávání obrazu, může vést i k náhodným chybám a nepředvídatelným výsledkům. Jedna studie ukazuje případy, ve kterých autorům stačí změnit pouhý jeden pixel na fotce a ANN ji poté interpretuje radikálně jinak. Jeden pixel jinak způsobí, že místo jelena ANN klasifikuje letadlo, nebo o obrázku koně s 99.9% jistotou tvrdí, že se jedná o žábu.<sup>38</sup> Změna naprostého detailu tak způsobí totálně protichůdné výsledky. Promyšleno do důsledků, jde o poměrně nebezpečný efekt – představa, že se podobný případ nastane například v kamerovém systému samořídícího auta (a to se netýká pouze zpracování obrazové informace, ale těchto modelů obecně), už přestává být groteskní.

Pravděpodobně nejde o pouhou výpočetní chybu, ale riziko interpretace obecně, kdy jeden špatně vyložený detail vede k řetězové reakci dalších zcestných úsudků a mylných přesvědčení, ve kterých se původní evidence ztratí a následná „kospirační teorie“ může mít fatální důsledky. Bezpochyby se na tom podílí efekt emergence a fenomén fraktální estetiky – nicméně jak tomu předcházet?

#### Avantgardní inspirace

Představy CNN charakterizuje prostorová deformace struktury vnímaného – rozklad na fragmenty, hromadění různých úhlů pohledu a syntéza partikulárních znaků, jež nějakým způsobem reprezentují celek. V některých případech to připomíná

---

<sup>38</sup> Viz studie One pixel attack

působivou avantgardní montáž (saxofon) nebo hravou pestrobarevnou impresi (květiny).



Obrazy amerického umělce a inženýra od Googlu Mike Tyky, který spolupracoval na projektu DeepDream.<sup>39</sup> Vlevo je obraz *Saxophone dreams* a vpravo obraz *Bacchus*. Ten první byl vytvořen tak, že se CNN zadalo vyvolat koncept saxofonu, ten druhý tak, že se nechalo CNN vyvolat koncept džbánu a květin zároveň.

Ve všech případech to nicméně něco o perspektivě ANN vypovídá. Namísto prostorového uspořádání a systematizace dominují znaky – představy nemají žádnou striktní kompozici a skládají se z kubistických fragmentů kombinujících různé perspektivy a různá zobrazení též detailů. Možná bychom očekávali, že hlavním vodítkem pro rozpoznání saxofonu bude jeho charakteristický tvar, kontury či silueta, aby se tak odlišil od ostatních jazzových nástrojů – ANN však zachycuje spíše celkovou „náladu“ nebo „zkušenost“ saxofonu – vykresluje složitou spleť potrubí, klapek a náústků s tvářemi muzikantů, metalický lesk mosazného plechu nebo hněd kabátů jazzových večerů.

Celek tak není identifikován skrze jeho uspořádání – pro ANN nejsou důležité proporce, délky a míry, není ani důležité jak jsou jednotlivé komponenty uspořádány – kde se nachází hubice, eso, tělo nebo zvon, nebo jestli to do sebe pasuje. Není

<sup>39</sup> <http://www.miketyka.com/> (Přístup duben 2019)

důležitá instrumentalita ani přítomnost funkčních částí, jak bychom očekávali i v prvních dětských kresbách. V tomto případě nicméně nejde o rafinovanou uměleckou stylizaci a překonání konceptuálního realismu, ale pravděpodobně o jistý ryze vizuální způsob vnímání, který není kontaminován zkušeností prostoru, nezná hloubku, neodděluje popředí od pozadí a detail od celku (na obraze *Saxophone dreams* ANN zobrazuje kromě saxofonu jako důležitý atribut také hráče na saxofon, téměř jako by byl součástí instrumentu), ale o to pečlivěji zkoumá povrch a texturu, prostorovost vynahrazuje množstvím perspektiv a úhlů pohledu, čímž je možná paradoxně na cestě k podobným cílům, o které usilovala avantgarda – oprostit se od strojených forem a zaběhlých kategorií, směřovat k subjektivnímu, niternému, podvědomému – dobrat se tak přirozenějších a spontánnějších způsobů vnímání, skrze které lze odhalit skryté významy všedních věcí.

–

Viděli jsme, že hranice imaginace CNN ještě zdaleka nebyly prozkoumány.<sup>40</sup> Přesto nabízejí nespočet podnětů k dalším analýzám, z nichž některé jsem se pokusil naznačit. Vyplývá z toho, že estetiku inceptionismu, která se zpočátku zdá bizarní a nepřístupná, lze interpretovat skrze široké množství již popsanych fenoménů – od analýzy ornamentů přes fraktální estetiku, nebo dokonce některé avantgardní postupy. Zajímavé by zřejmě bylo metodu také obrátit a na základě CNN zkoumat rozdíly a sjednocující prvky jednotlivých estetik. Primárně to však naznačuje některé podstatné rysy vizuálního vnímání jako takového – zobrazení psů a pagod evidentně není pouze záležitostí kvantity podobných obrázků v datasetu, nýbrž jistých univerzálních estetických kritérií. U člověka jich je charakterizováno několik: například preference symetrie, zlatého řezu, fraktálů nebo opakujících se vzorů.<sup>41</sup> A

---

<sup>40</sup> Respektive pokud se vůbec začaly zkoumat. Reakce společnosti na DeepDream mě prostě udivuje – člověk by očekával, že zveřejnění klíčových článků v roce 2015 otřese společností a filosofickým diskurzem, začnou se klást podnětné otázky, spustí se lavina interpretací, analýz a mezioborových diskusí, nebo že to alespoň bude působit provokativně – místo toho se téma vyčerpalo několika senzativními články v médiích a trendními hashtagy na twitteru, kde lidé sdíleli fotografie transformované do podoby psích obličejů podle inceptionistické estetiky (ovšem bez jakéhokoliv originálního přínosu), což se záhy okoukalo a utichlo. Tento postoj vůči výzkumu v oblasti umělé inteligence mi přijde zvrácený, pokud ne přímo nebezpečný.

<sup>41</sup>Viz *Cross-cultural empirical aesthetics*, 2018

<https://www.sciencedirect.com/science/article/pii/S0079612318300025> (Přístup duben 2019) # How to



zdá se, že se to s CNN do velké míry shoduje – především pak po stránce druhých dvou zmiňovaných kritérií. Co se týče vnímání na nižších úrovních abstrakce, citlivost CNN na textury a ornamenty je mimořádně vysoká – zatímco pro člověka je přirozený konceptuální realismus, pro CNN spíše impresionismus – to potvrzují i výsledky některých studií ukazujících, že zatímco člověk dokáže obraz lépe než CNN rozpoznat na základě jeho kontur nebo siluety, CNN naopak vítězí v oblasti textury.<sup>42</sup> Jde o jistě zajímavý aspekt z výtvarného hlediska, jelikož síť tak vnímá obraz ryze vizuálně, bez jakéhokoliv předporozumění, jaké spatřujeme třeba v dětské kresbě – proto se některé výtvary CNN spíše než jim podobají modernistickým malbám.

Kromě toho bychom mohli tento typ vidění analyzovat na základě spojení s vizuální kúrou savců – pečlivé srovnání by mohlo vyjevit například větší souvislosti s vnímáním konkrétních zvířat – a dávalo by smysl, kdyby některá zvířata byla také výrazně citlivější na textury a s tím i pohyb a dynamiku obrazu.<sup>43</sup> Nicméně přes uvedené výhody jsou dobré důvody, proč usilovat o to, aby CNN dokázala vnímat na základě abstraktnějších kritérií srovnatelných s člověkem – například přítomnost nepředvídatelných chyb (viz Groteskní omyly aneb Efekt motýlích křídel) nebo neschopnost rozlišovat na obrazech pozadí od popředí a objekt od kontextu. Zatím se to ve většině případů řeší kvantitou dat, z filozofie vnímání by však vyplývalo, že k plnohodnotné vizuální zkušenosti je nutně zapotřebí kategorie času, prostoru, nebo kontextu více smyslových vjemů – domnívám se, že teprve s určitou mírou této synestezie bude vnímání CNN abstraktnější, spolehlivější a možná bližší tomu lidskému.

---

Explore Golden Ratio in Architecture and Designing City, 2016  
[https://www.researchgate.net/publication/308432227\\_How\\_toExplore\\_Golden\\_Ratio\\_in\\_Architecture\\_and\\_Designing\\_City](https://www.researchgate.net/publication/308432227_How_toExplore_Golden_Ratio_in_Architecture_and_Designing_City) (Přístup duben 2019) # Why golden rectangle is used so often by architects: A mathematical approach, 2015 <https://www.sciencedirect.com/science/article/pii/S1110016815000265> (Přístup duben 2019) # List of works designed with the golden ratio, 2018 [https://en.wikipedia.org/wiki/List\\_of\\_works\\_designed\\_with\\_the\\_golden\\_ratio](https://en.wikipedia.org/wiki/List_of_works_designed_with_the_golden_ratio) (Přístup duben 2019) # African Fractals: Modern Computing and Indigenous Design by Ron Eglash, 1999 [https://monoskop.org/images/f/fc/Eglash\\_Ron\\_African\\_Fractals\\_Modern\\_Computing\\_and\\_Indigenous\\_Design.pdf](https://monoskop.org/images/f/fc/Eglash_Ron_African_Fractals_Modern_Computing_and_Indigenous_Design.pdf) (Přístup duben 2019)

<sup>42</sup> viz Geirhos <https://openreview.net/forum?id=Bygh9j09KX> (Přístup duben 2019)

<sup>43</sup> Pokud by se takové souvislosti prokázaly, byla by to rovněž unikátní příležitost jak „vidět jejich očima“.

I takto nicméně CNN nabízí zajímavou simulaci procesů, které se pravděpodobně odehrávají i v lidské vizuální kůře. CNN tak lze interpretovat také jako analogii podvědomí – v analýze snů a halucinací bychom tak mohli dojít k překvapivým závěrům. Například k tomu, že jsou překvapivě bohaté i přes absenci jakékoliv intence, skryté tužby, sexuality nebo agrese (viz Freud). Opačná interpretace by naopak ukazovala, že vizuální percepce zaujímá v podvědomí výsadní postavení a intence jsou až jakousi její nadstavbou. Merleau-Pontyho tvrzení o primátu vnímání by tak platilo i zde.

Je vidět, že CNN a metoda inceptionismu tak poskytují podnětný přístup k problematice ANN a umělé inteligence celkově, přičemž skrze estetickou metodu lze pokládat zajímavé otázky, které mají mezioborový význam.

## Seznam literatury

Není-li uvedeno jinak, citace ze zdrojů v angličtině jsou vlastního překladu

1. Floridi, Luciano, *Philosophy and Computing*, Routledge, New York, 1999, elektronická verze
2. Garson, James, "Connectionism", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/fall2018/entries/connectionism/>
3. Spehar, B., Clifford, C. W. G., Newell, B. R., & Taylor, R. P. (2003). *Universal aesthetic of fractals*. *Computers & Graphics*, 27(5), 813–820. doi:10.1016/s0097-8493(03)00154-7
4. Merleau-Ponty, M., *Primát vnímání a jeho filosofické důsledky*, přel. J. Halák, Togga, Praha 2011 (dále citováno jako: *Primát vnímání*).
5. Merleau-Ponty, M., *Okno a duch a jiné eseje*, Obelisk, Praha 1971