

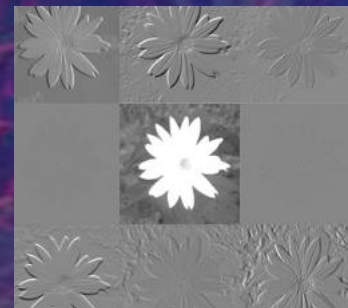
Průzkum sítě

a generování obrazů

Black box

- Neuronové sítě jsou, na rozdíl od klasických algoritmů strojového učení, uzavřené systémy. Nikoliv uzavřené, že nevíme co v nich je (natrénované hodnoty vidíme), ale uzavřené, že nerozumíme tomu co se ty hodnoty znamenají. Kdyby hodnoty v neuronech byly úplně jiné, a klasifikovala by pořád stejně, nebude to vůbec podezřelé.
- Není pak zcela zřetelné, jak se síť bude chovat v krajních případech. Jde také snadno obalamutit.
- Existují způsoby jak se podívat síti „na zoubek“

Aktivace



- „Co vidí neuronka, když do ní pošleme vstupní data.“
 - Ani ty data posílat nemusíme. Aktivační funkce (activations) požere i větší obrázek než je vstupní velikost cnn. Jednoduše vyzobne ze sítě konvoluční filtry a ty uplatní na jakýkoliv obrázek.
- (Pozadí této prezentace jsou upravené aktivace alexnet kytkou256)
- (přečtete si nejdříve následující slajd) a vytvořte aktivace z třetí konvoluční vrstvy sítě.
- Obecně konvoluční vrstvy na začátku sítě detekují jednoduché vlastnosti jako jsou barvy a hrany. Můžete se o tom přesvědčit tím, že po zjištění aktivací z první konvoluční vrstvy je žlutá kytko celá aktivovaná (je bílá). V pozdějších konvolučních vrstvách se vlastnosti (features) kombinují ve složitější. Což je hezky (lépe než na kytce) vidět [zde](#), kdy si neuronka natrénovala filtry pro detekci oka (a tím pádem i obličeje atd...).

Průzkum sítě

- Načtěte síť squeezenet (stejným příkazem) a analyzujte ji pomocí `analyzeNetwork`. Ve třetím sloupečku máte aktivace, tedy jaký objekt (o jakých rozměrech) získáte, pokud sáhnete k aktivacím z dané vrstvy. Learnables jsou hodnoty, které se mění s učením sítě. U konvolučních vrstev jsou to jádra konvolučních filtrů. To co se učí konvoluční neuronka jsou tedy především hodnoty těchto filtrů (a potom plně propojené vrstvy na konci sítě).

ANALYSIS RESULT				
	Name	Type	Activations	Learnables
1	data 227x227x3 images with 'zerocenter' normalization	Image Input	227x227x3	-
2	conv1 64 3x3x3 convolutions with stride [2 2] and padding [0 0 0]	Convolution	113x113x64	Weights 3x3x3x64 Bias 1x1x64
3	relu_conv1 ReLU	ReLU	113x113x64	-
4	pool1 3x3 max pooling with stride [2 2] and padding [0 0 0]	Max Pooling	56x56x64	-
5	fire2-squeeze1x1 16 1x1x64 convolutions with stride [1 1] and padding [0 0 0]	Convolution	56x56x16	Weights 1x1x64x16 Bias 1x1x16
6	fire2-relu_squeeze1x1 ReLU	ReLU	56x56x16	-
7	fire2-expand1x1 64 1x1x16 convolutions with stride [1 1] and padding [0 0 0]	Convolution	56x56x64	Weights 1x1x16x64 Bias 1x1x64

Adversarial images

- (nepřátelský obraz)
- Cíleně vytvořený obraz, tak aby se člověk a neuronka se naprosto rozcházeli v klasifikaci.
 - „Pro člověka vypadá jako květina, pro síť jakožto plameňák“
- Klasifikujte obrázek (what_is_it.png). Vypadá dosti jako náhodný šum.
 - Podobá se něčím obrázek třídě do které je klasifikování?
- Zkombinujte obrázek kytky a zdánlivého šumu, tak aby člověk klasifikoval jako kytku, síť však nikoliv. (předpokládejte, že správná třída pro kytku je daisy).

Vizualizace



- deepDreamImage – generování obrázku pomocí CNN.
 - „Jak neuronka vidí danou třídu“
 - Iterativní proces, kdy neuronka upravuje obraz tak aby jí co nejvíce aktivoval.
 - Vznikají z toho prapodivné obrazy se skutečnými vlastnostmi původní třídy.
 - Zde obrázek když alexnet sní o Daisy
 - V matlabu je pravděpodobně jakási chyba v implementaci a tak jsou obrázky takové zašedlé. Proto [toto](#).
 - Můžete díky tomu pochopit na co neuronka cílí.
- Proč se neuronka rozhodla takto pro klasifikaci..?
<https://blogs.mathworks.com/deep-learning/2019/01/31/deep-learning-visualizations-cam-visualization/>

Další generování obrázků

- Přenos stylu -> Přenést styl jedné male na jiné obrázky (pak to třeba vypadá jako by je nakreslil jiný malíř).
- Generative Adversarial Network (GAN) -> Způsob generování obrazů. Je potřeba 2 sítě a databázi obrazů jejichž podoby chcete generovat (například obličeje). Jedna neuronka generuje obrázky a předhazuje je té druhé, ta odhaduje jestli jí byl předhozen generovaný nebo obrázek z databáze. Obě neuronky se neustále doučují a tím pádem navzájem vylepšují, až generované obrázky vypadají uvěřitelně. Viz: <https://www.thispersondoesnotexist.com/> (skutečně nikdo z těchto lidí neexistuje, a občas jsou tam děsivé artefakty)