

Párhuzamos képstílus átruházás konvolúciós neuronhálókkal

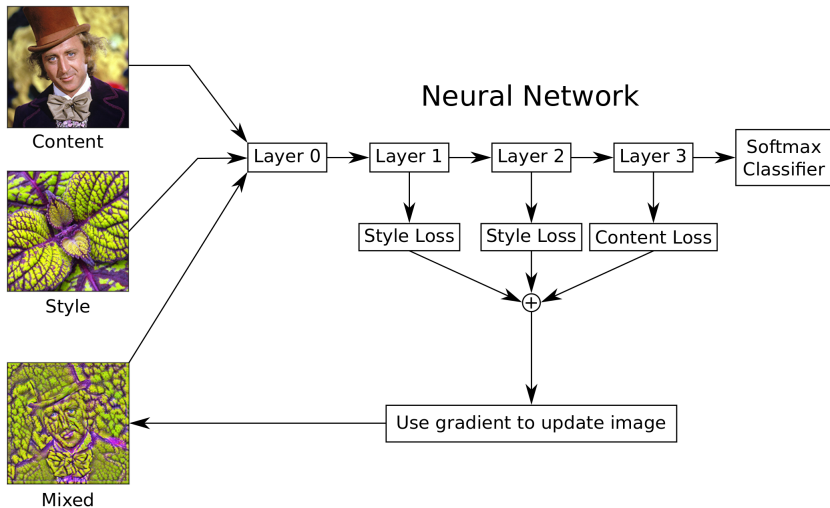
Szilágyi Ervin

Témavezető: Dr. Iclanzan Dávid

Sapientia Erdélyi Magyar Tudományegyetem
Műszaki és Humántudományok kar
Szoftverfejlesztés szak

2017. július 2.

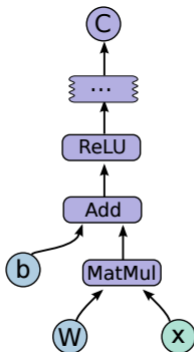
Mi értünk stílusátruházás alatt?



A dolgozat célja

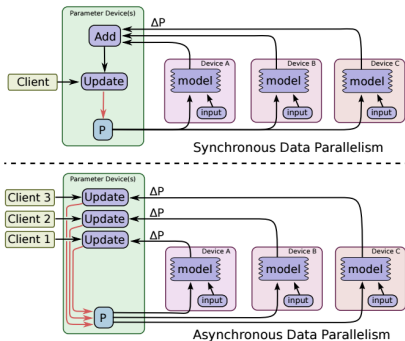
- Grafikus felhasználói felülettel rendelkező alkalmazás fejlesztése
- Gépi tanulást (deep learning) alkalmazó rendszer tervezése és megvalósítása
- Híres magyar festők ismertebb műveinek művészeti stílusát átruházni képekre / mozgóképekre
- Párhuzamos gépi tanítási folyamat ami kihasználja a GPU által biztosított párhuzamosítási lehetőségeket

A Tensorflow könyvtár bemutatása

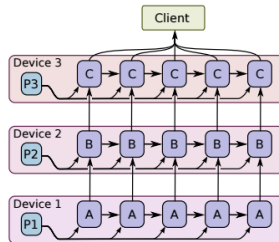


ábra. Tensorflow számítási gráf

Párhuzamos tanítás a Tensorflow segítségével



(a) Adatpárhuzamos megközelítés



(b) Feladatpárhuzamos megközelítés

ábra. Párhuzamos tanítás

- "Deep learning" tanítási metódus
- Előre betanított neuronháló (VGG19: 16 konvolúciós réteg, 5 pooling réteg)
- Statikus kép esetében külön tanításra kerül a bemeneti kép és a stílus kép is
- Mozgókép esetében minden képkocka tanításra kerül
- Temporális összefüggések a képkockák között

A tanításhoz használt neuronháló

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

ábra. VGG-19 háló szerkezete

A bemeneti kép tanítása

- A konvolúciós szűrők tartalmazzák a kép sajátosságait
- Egy adott betanított réteg válasza egy bemeneti képre vizualizálható, ha fehér zaj képre értékeljük ki azt

A bemeneti kép veszteségfüggvénye felírható mint:

$$L_{content}(\vec{x}, \vec{r}, l) = \frac{1}{2} \sum R_{ij}^l - W_{ij}^l \quad (1)$$

Ahol:

- \vec{x} - a bemeneti képet
- R^l - az l -edik réteg válasza a bemeneti képre
- W^l - az l -edik réteg válasza a fehér zaj bemenetre
- \vec{r} - pedig azt a kimeneti képet jelenti amit a rendszer generál a rétegek tulajdonságaiból

Kiértékelt réteg: *conv4_2*

A stílus kép tanítása

A Gramm-matrix ismertetése

- A Gramm mátrix egy szorzatot jelen egy adott vektorhalmaz összes elemei között.
- Hogyha adott egy vektorhalmazunk $v_1 \dots v_n$, akkor a G Gramm mátrixot a következő eljárás szerint határozzuk meg:

$$G_{ij} = v_i \cdot v_j \quad (2)$$

- A Gramm mátrix ij pozíciójában elhelyezkedő elem megadja, hogy egy adott réteg i -dik tulajdonsága mennyire teljesül a j -dik tulajdonság jelenlétében,

A stílus kép tanítása

A veszteségfüggvény meghatározása

Ha az l rétegnek N szűrője van, akkor felírható $G \in R^{N_l \times N_l}$ Gramm-mátrix, ahol:

$$G_{ij}^l = \sum_k F_{ik}^l \cdot F_{jk}^l \quad (3)$$

A veszteségfüggvény egyetlen rétegre felírható mint a fehér zaj kép Gramm mátrixa és a stílus kép Gramm mátrixának átlagos négyzetes hibájaként:

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij} - A_{lj})^2 \quad (4)$$

A stilizált kép tisztítása

Total Variation Denoising

A stilizált képet és eltoljuk X koordináta mentén egy pixellel, majd az Y koordináta mentén is eltoljuk egy pixellel.

$$L_{tv}(\vec{a}, \vec{x}) = \sum_{i,j} |(X_{ij} - A_{i+1j})| + \sum_{i,j} |(X_{ij} - A_{ij+1})| \quad (5)$$

A rendszer tesztelese

