

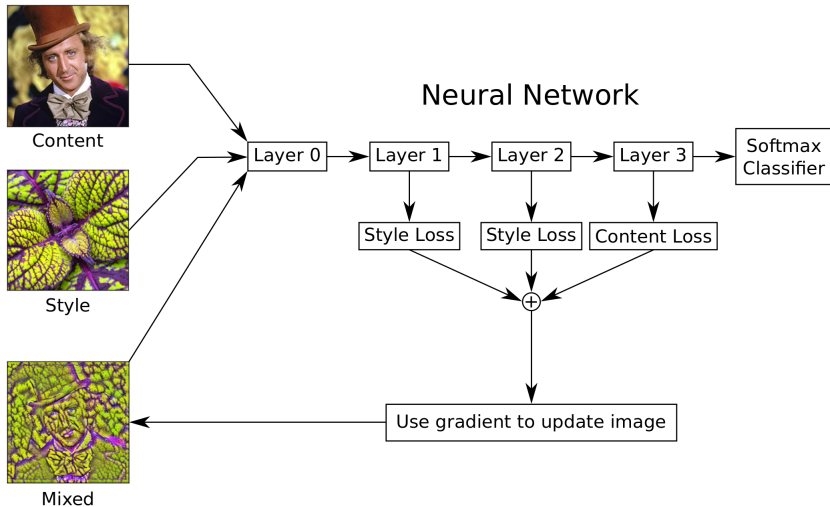
# Párhuzamos képstílus átruházás konvolúciós neuronhálókkal

Szerző: Szilágyi Ervin  
Témavezető: Dr. Iclănzan Dávid

Sapientia Erdélyi Magyar Tudományegyetem  
Műszaki és Humántudományok kar  
Szoftverfejlesztés szak

2017. július 2.

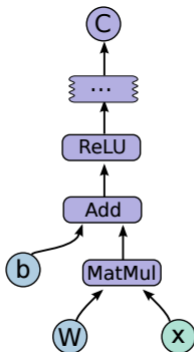
# Miértünk stílusát ruházás alatt?



# 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

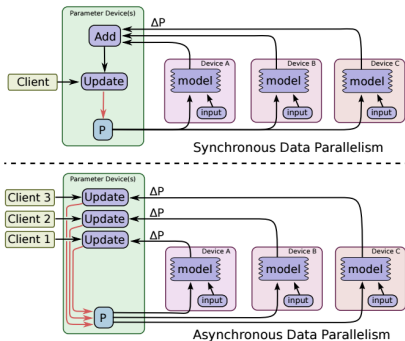
# A Tensorflow könyvtár bemutatása



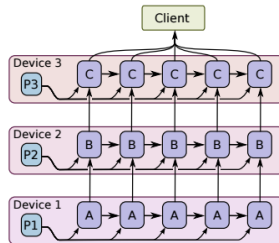
ábra. Tensorflow számítási gráf

Forrás: *Google.com*

# Párhuzamos tanítás a Tensorflow segítségével



(a) Adatpárhuzamos megközelítés



(b) Feladatpárhuzamos megközelítés

ábra. Párhuzamos tanítás

Forrás: Google.com

- "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
- Mozgóképek 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 <b>LRN</b>	conv3-64 <b>conv3-64</b>	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 <b>conv3-128</b>	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 <b>conv1-256</b>	conv3-256 conv3-256 <b>conv3-256</b>	conv3-256 conv3-256 conv3-256 <b>conv3-256</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 <b>conv1-512</b>	conv3-512 conv3-512 <b>conv3-512</b>	conv3-512 conv3-512 conv3-512 <b>conv3-512</b>
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.

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

# Mozgóképf tanítása

## Naív megközelítés

A videót feldaraboljuk képkockákra, majd az összes képkockára átruházzuk a stílust.

Előnyei:

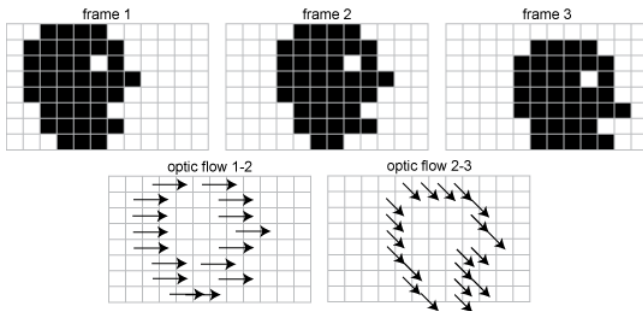
- Gyors.

Hátrányai:

- Nincs folyamatos átmenet a képkockák között
- Artifacts, pop-ins

# Mozgóképfé tanítása

## Optical flow bevezetése

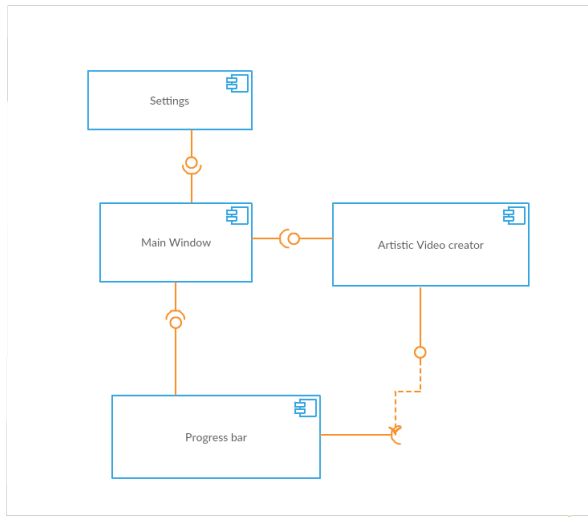


Forrás: [http://www.scholarpedia.org/article/Optic\\_flow](http://www.scholarpedia.org/article/Optic_flow)

ábra. Optical flow szemléltetése

# A szoftvet architektúrája

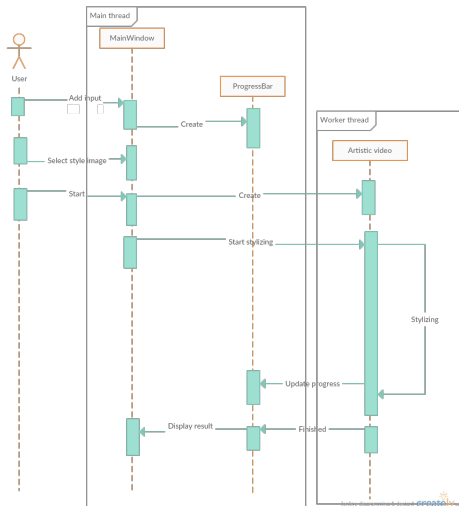
## A szoftver komponensei



ábra. A szoftver komponensei

# A szoftvet architektúra

## Többszálas megvalósítás



# A rendszer tesztelése

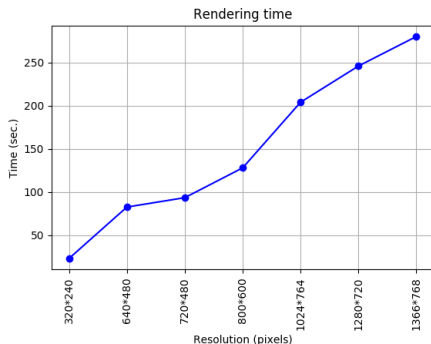
## A tesztkonfiguráció meghatározása

- **Alaplap:** MSI Z170A-G45 GAMING
- **Processzor:** Intel(R) Core(TM) Skylake i7-6700 ("non-K") CPU @ 3.40GHz (Turbo Boost: 3.90GHz)
- **Videokártya:** GIGABYTE GeForce GTX 1080 G1 GAMING 8GB DDR5X 256-bit
- **RAM Memória:** Corsair Vengeance LPX Black 32GB DDR4 3000MHz
- **Merev lemez:** HyperX Savage SSD, 240GB, 2.5", SATA III
- **Operációs rendszer:** Ubuntu 17.04 LTS

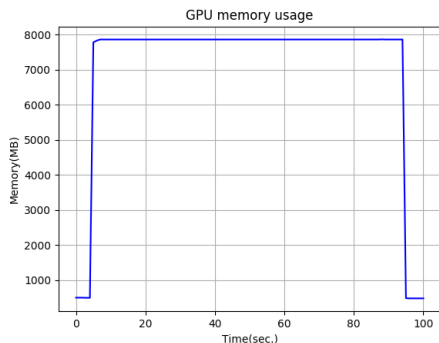


# Egy képkockára történő stílusátruházási idő

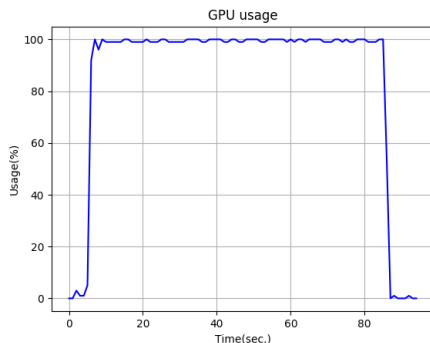
Felbontás(pixel)	Idő(másodperc)
320×240	23.126104
640×480	82.640347
720×480	93.487573
800×600	127.973041
1024×764	203.892024
1280×720	245.769916
1366×768	279.945757



# Videókártya memóriahasználata és kihasználtsága



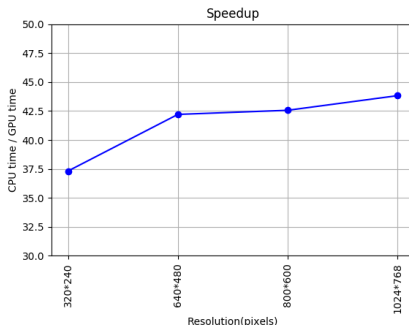
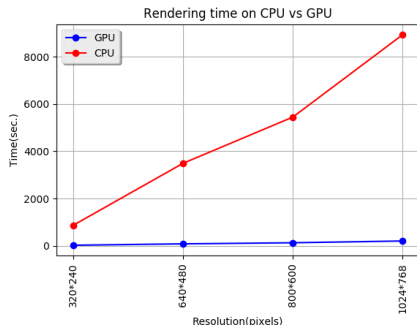
ábra. Memóriahasználat



ábra. Kihasználtság

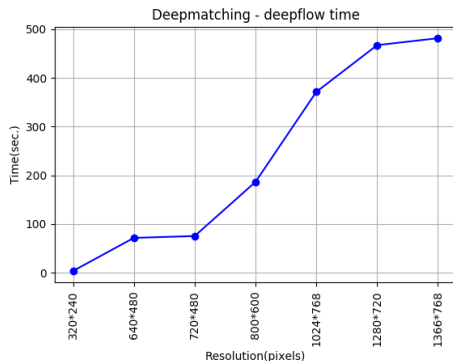
# CPU - GPU összehasonlítás

Felbontás(pixel)	GPU Idő(sec.)	CPU Idő(sec.)	Gyorsulás
320×240	23.126104	862.980290	37.316285
640×480	82.640347	3487.931886	42.206162
800×600	127.973041	5446.849902	42.562479
1024×768	203.892024	8935.804122	43.826158



# Az Optical flow időigénye

Felbontás(pixel)	Idő(másodperc)
320×240	3.700949
640×480	71.693549
720×480	75.388477
800×600	186.4160516
1024×764	371.140060
1280×720	466.996901
1366×768	481.25628

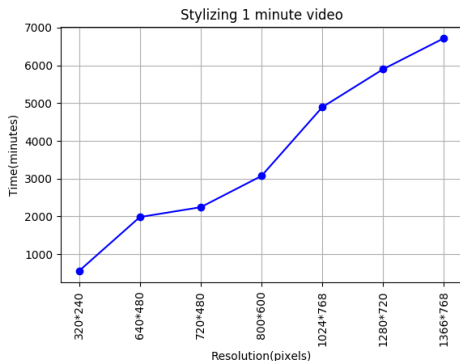


# Egy perces videóra történő stílusátvitel

24 FPS, 60 másodperc, 1440 képkocka



ábra. Optical flow használatával



ábra. Optical flow használata nélkül

- Grafikus felhasználói felülettel rendelkező szoftver fejlesztése
- Magyar festők híres műveinek stíluást alkalmazni mindennapi képekre/mozgóképekre
- Képkockák közötti temporális összefüggések kihasználása mozgóképek esetében
- Tesztek, mérések elvégzése

Továbbfejlesztési lehetőségek:

- Saját háló betanítása, más tanítási eljárás használata
- Optical flow algoritmus gyorsítása (párhuzamosítás, más algoritmusok kipróbálása)
- Teljes Windows-os támogatás
- Grafikus felület újratervezése, szépítése
- Cloud alapú szolgáltatás készítése

Köszönöm a figyelmet!