

# Video Resolution Upscaling Using Neural Networks

Yana Zabrodszkaya

ČVUT - FIT

zabroyan@fit.cvut.cz

22. prosince 2020

## 1 Úvod

Úkolem semestrální práce bylo vzít krátké video nízkého rozlišení, pomocí GANu<sup>1</sup> a autoencoderu vygenerovat video vyššího rozlišení a porovnat výsledky.

Jako video jsem zvolila scénu z poslední bitvy z anime *Naruto: Shippuuden*. Ačkoli epizoda, na které jsem pracovala, je z roku 2017, špatná kvalita videa je u starších anime častým problémem.

## 2 Vstupní data

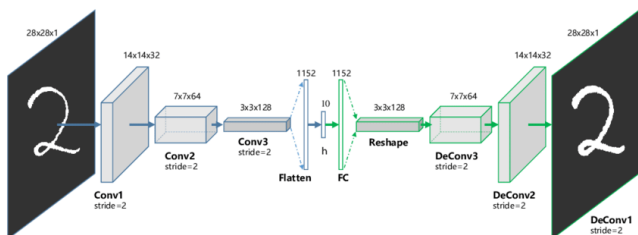
Používala jsem dvě videa stejného obsahu: jedno s rozlišením 240p, druhé – 480p. Velikost obou videí byla změněna na  $480 \times 852$  pixelů. Z každého videa bylo extrahováno 681 snímků. Video s nižším rozlišením bylo použito za účelem trénování modelů, s vyšším – pro porovnání výsledků.

## 3 Metody

### 3.1 Autoencoder

Autoencoder je neuronová síť, která se učí, jak zredukovat data a zrekonstruovat je tak, aby byla co nejvíc podobná původním datům.

Autoencoder se skládá z encoderu a decoderu. Cílem encoderu je komprimovat data do nižší dimenze takovým způsobem, že přetrvávají pouze nejvýznamnější vlastnosti, a cílem decoderu je zrekonstruovat data takovým způsobem, aby co nejvíc odpovídala originálu[3].



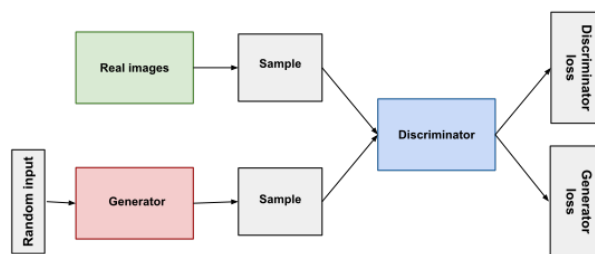
Obrázek 1: Příklad autoencoderu[2]

Implementace autoencoderu byla inspirována článkem[3] a upravena tak, aby odpovídala počí-

tačové výkonnosti. Výsledný autoencoder má něco přes 1 milion parametrů a skládá se z 17 vrstev.

### 3.2 GAN

GAN se skládá ze dvou sítí: generátoru, který se učí generovat data, a diskriminátoru, který se učí rozpoznávat falešná data.



Obrázek 2: Příklad GANu

Implementace GANu byla inspirována článkem[1] a upravena tak, aby odpovídala počítačové výkonnosti. Ve výsledku generátor má 366 tisíc parametrů a 32 vrstvy a diskriminátor má 6 milionů parametrů a 10 vrstev.

## 4 Výsledky

Příklad testovacího snímku (s rozlišením 480p) je ukázán na obrázku 3.



Obrázek 3: Originál (480p)

Příklad trénovacího snímku (s rozlišením 240p) je ukázán na obrázku 4.

Každý model vygeneroval 681 snímků, které následně byly sloučeny do videa.

<sup>1</sup>Generative Adversarial Network

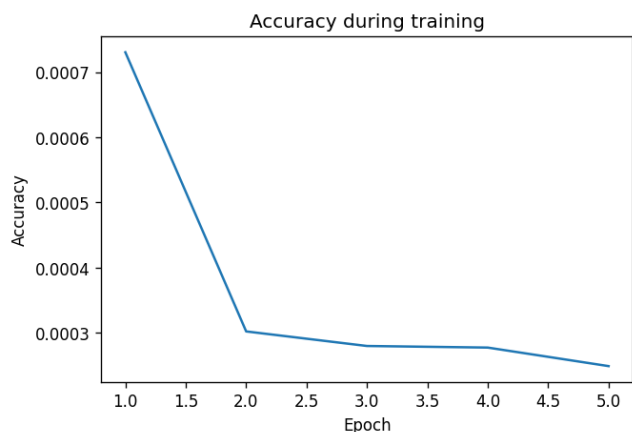


Obrázek 4: Originál (240p)

#### 4.1 Autoencoder

Autoencoder se učil mnohem lépe a už od začátku měl přesnost na testovacích datech 91%. Po pěti epochách dosahl přesnosti 97% (obrázek 5).

Výsledný snímek je vidět na obrázku 6. Kvalita je lepší než u 240p, ale kvůli malému počtu epoch nedosahuje požadované kvality 480p.



Obrázek 5: Přesnost u autoencoderu

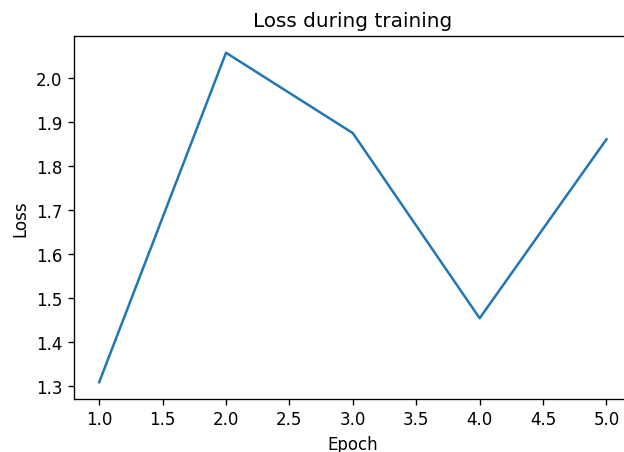


Obrázek 6: Výsledek (autoencoder)

#### 4.2 GAN

GAN dosahl horších výsledků než autoencoder. Změna ztrátové funkce modelu je na obrázku 7.

Snímek vygenerovaný po pěti epochách je ukázán na obrázku 8 a obsahuje hodně šumu.



Obrázek 7: Ztráta u GANu



Obrázek 8: Výsledek (GAN)

### 5 Závěr

Táto semestrální práce byla výbornou příležitostí ke procvičení neuronových sítí.

Bohužel, jsem nedosáhla ideálních výsledků kvůli počítačové výkonnosti. Experimentálně jsem zjistila, že maximální velikost batche je 3 a maximální počet snímků pro trénování je kolem 300, aby to nedocházelo k nedostatku paměti nebo restartování jádra.

Předpokládám, že s lepším modelem a delším trénováním lze výsledný model použít ke zlepšení kvality jakéhokoli animovaného videa.

### Reference

- [1] Deepak Birla. Single image super resolution using gans — keras. online, 2018. [cit. 2020–12–22] [shorturl.at/fAPR9](https://shorturl.at/fAPR9).
- [2] Kapil Chauhan. Super-resolution using autoencoders and tf2.0. online, 2020. [cit. 2020–12–22] [shorturl.at/iMS78](https://shorturl.at/iMS78).
- [3] Harshil Patel. Image super-resolution using convolution neural networks and auto-encoders. online, 2020. [cit. 2020–12–22] [shorturl.at/bg1HZ](https://shorturl.at/bg1HZ).