Seeing music using deepsing: Creating machine-generated visual stories of songs

Nikolaos Passalis[1] and Stavros Doropoulos[2]

[1]Postdoctoral Researcher, AUTh
[2]CIO, DataScouting

Futurama Holophonor

What is Music?

- **What is music?**

What is Music?

- **What is music?**
  - *Music is an art form and cultural activity whose medium is sound organized in time.*
- **What is music for you?**

What is Music?

- **What is music?**
  - *Music is an art form and cultural activity whose medium is sound organized in time.*
- **What is music for you?**
  - “music to my ears”,
  - ..., pleasant experience, ...
  - ..., travel, ...
  - ..., imagination, ...
  - ..., feelings, ...

What is Music?

- **What is music?**
  - *Music is an art form and cultural activity whose medium is sound organized in time.*
- **What is music for you?**
  - "music to my ears",
  - ..., pleasant experience, ...
  - ..., travel, ...
  - ..., imagination, ...
  - ..., feelings, ...
- A way of communicating **feelings**!

Why music?

- Music is a universal way of communicating
- At the same time, music has a limited (to nonexistent) capacity of transferring semantic information
- However, it excels at transferring emotions

Why music?

- Music is a universal way of communicating
- At the same time, music has a limited (to nonexistent) capacity of transferring semantic information
- However, it excels at transferring emotions
- What do you feel?

Why music?

- Music is a universal way of communicating
- At the same time, music has a limited (to nonexistent) capacity of transferring semantic information
- However, it excels at transferring emotions
- What do you feel?

- Major scales are generally described as “happy”, while minor ones as “sad”

Music and Imagination

- Music is capable of triggering our imagination
- Several people describe music as a way to *travel*
- Listen to this and close your eyes...

Music and Imagination

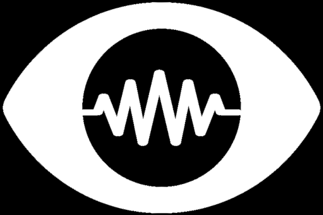
- Music is capable of triggering our imagination
- Several people describe music as a way to *travel*
- Listen to this and close your eyes...
- What do you feel?
- Did you see anything?

Music and Imagination

- Humans tend to **mentally visualize music**!
- The question that has been bothering us for more than 2 years...
- **Are machines capable of imagining and dreaming concepts when listening to music?**

Music and Imagination

- Humans tend to **mentally visualize music**!
- The question that has been bothering us for more than 2 years...
- **Are machines capable of imagining and dreaming concepts when listening to music?**
- Now, look at this ...



deepsing.com

Music to Image Translation

- deepsing is a **deep learning method** for performing attributed-based music-to-image translation
- deepsing works by **synthesizing visual stories according to the sentiment expressed by songs**
- The generated images aim to induce the same feelings to the viewers, as the original song does, reinforcing the primary aim of music, i.e., communicating feelings

Music to Image Translation

- But how this works?

Music to Image Translation

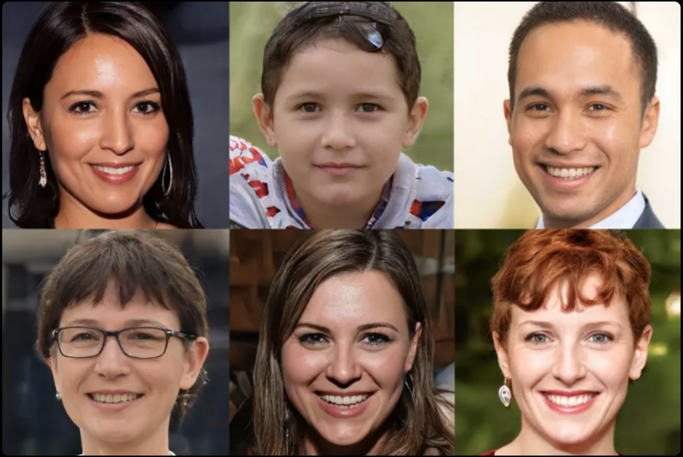
- But how this works?
- Let first revisit Deep Learning...
    - Neural Networks are capable of extracting the sentiment (valence and arousal) from audio segments
    - Generative Adversarial Networks (GANs) can generate images by generalizing the knolwedge they have encoded

Generative Adversarial Networks

- How well do they work?

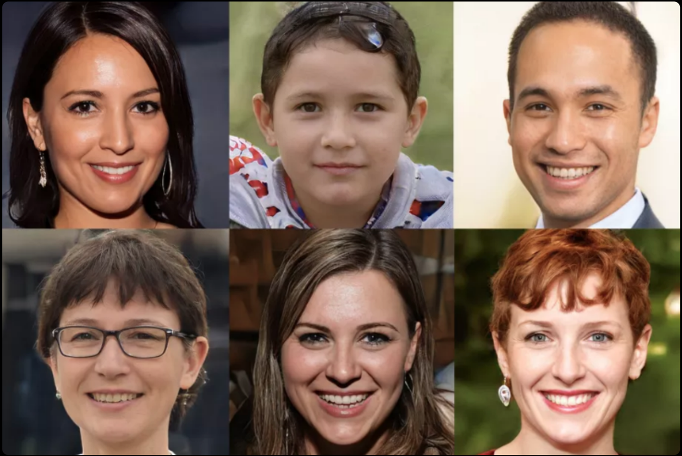
Generative Adversarial Networks

- How well do they work?



Generative Adversarial Networks

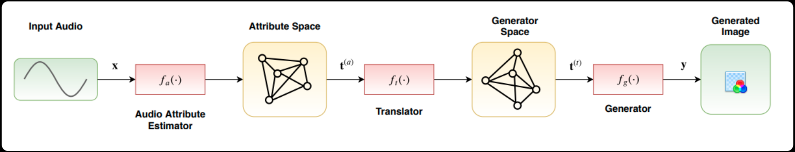
- How well do they work?



- None of these persons are real!

Music to Image Translation

- But how this works?



- How do we do the actual translation?

Music to Image Translation

- Approach 1: **Dictionary-based Translation**
  - Use a sentiment-dictionary to lookup the sentiment of each class
  - Any issues?

Music to Image Translation

- Approach 1: **Dictionary-based Translation**
  - Use a sentiment-dictionary to lookup the sentiment of each class
  - Any issues?
  - No guaranty that the image generated using the GAN for a specific class will indeed induce the same sentiment as the one given in a handcrafted dictionary

Music to Image Translation

- Approach 2: **Neural Translation**
  - Train a model to predict the sentiment of each image
  - Generate many GAN-based images
  - Learn how to **invert** the generation process in order to produce images with the correct sentiment
  - ... the model learns how to *dream* using a GAN!

Music to Image Translation

- Approach 2: **Neural Translation**
  - Train a model to predict the sentiment of each image
  - Generate many GAN-based images
  - Learn how to **invert** the generation process in order to produce images with the correct sentiment
  - ... the model learns how to *dream* using a GAN!

- Any issues?

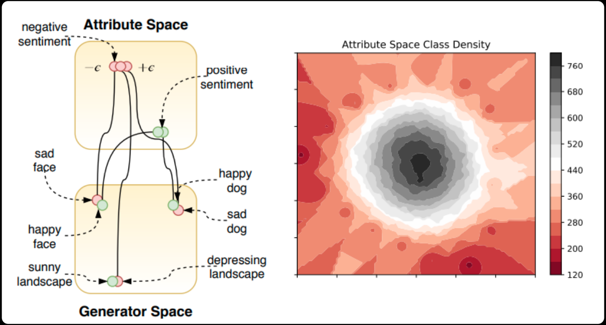
Music to Image Translation

- Approach 2: **Neural Translation**
  - Train a model to predict the sentiment of each image
  - Generate many GAN-based images
  - Learn how to **invert** the generation process in order to produce images with the correct sentiment
  - ... the model learns how to *dream* using a GAN!

- Any issues?
- Unfortunately yes...

Music to Image Translation

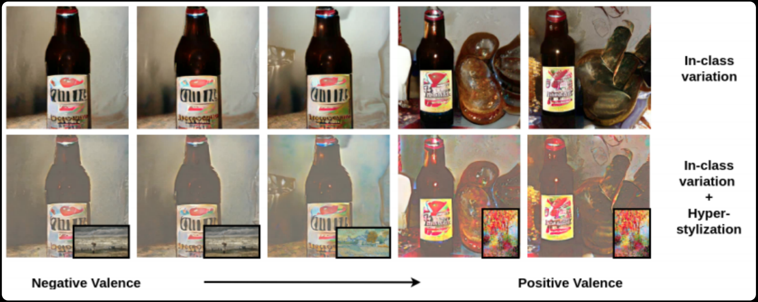
- There is no “1-1” between the used sentiment and semantic spaces leading to a chaotic mapping



- It was virtually impossible to train a translator without restricting the number of classes

Music to Image Translation

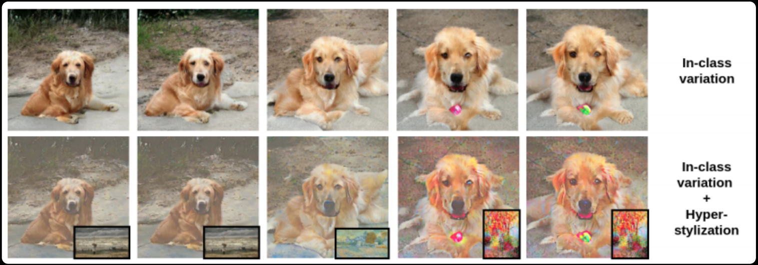
- **Our solution**: Perform density-based sampling on the classes
- Then, train the model using a small sub-sample of the classes
- This also allows for discovering different *sentiment views* for the same class:



- The sentiment can be further enhanced using Neural Style Transfer!

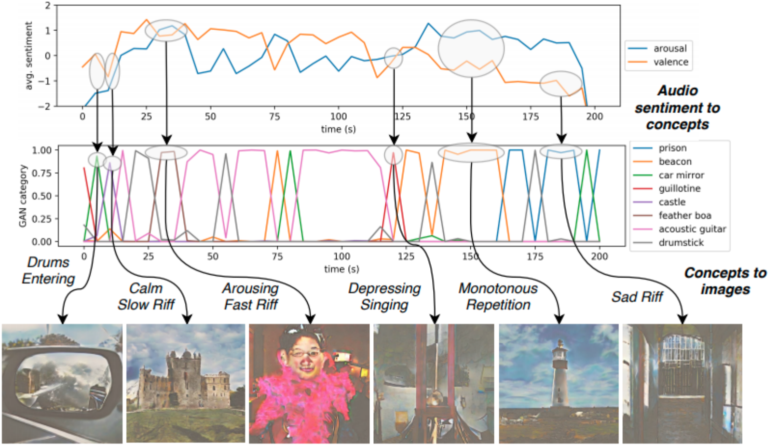
Music to Image Translation

- Our solution: Perform density-based sampling on the classes
- Then, train the model using a small sub-sample of the classes
- This also allows for discovering different *sentiment views* for the same class:



- The sentiment can be further enhanced using Neural Style Transfer!

Some examples...



Key frames selected along with annotations regarding the corresponding affective content of the song. For example, note the generated "feather boa" during the most arousing riff of the song and the transition to a "prison" as the valence of the song decreases. Sample frames generated using the song "Chop Suey!" by "System Of A Down".

Bringing well-known painters to life ...

- What if you could see how Van Gogh, Picasso, and others would paint when listening to a specific song?

Bringing well-known painters to life ...

- What if you could see how Van Gogh, Picasso, and others would paint when listening to a specific song?

deepsing can do!