BACKGROUND

Describe the customer's goal and plans

Users: Advertisers and video editors, individuals

Goals: Efficiently create original content for commercials and personal videos. Experiment with different styles without the high cost of hiring artists.

Pains: Struggling to stand out in the digital age, facing expensive artist fees, and looking for an affordable solution to personalize videos.

VALUE PROPOSITION

Propose the product with the value it creates and the pains it alleviates.

We propose ClipMorph, a tool to take your videos to the next level in a cheap, fast and reliable way. We will use AI to generate multiple stylized videos of your ad that will become much more compelling. The client will select a style or try multiple ones and we will bring his ideas to life!

Product: Platform that allows to upload videos and images and select a style to add

Alleviates: Add style to images/videos

Advantages: When the client uses our platform, he can directly see its stylized video, and can even try multiples styles to see what fits best.

OBJECTIVES

Breakdown the product into key objectives that need to be delivered.

1) Collect a catalog of styles to be offered for style transfer.
2) Apply a selected style among the catalog to a user-uploaded image/video.
3) Integrate a new style to the catalog, from a user-uploaded reference style image.

SOLUTION

Define the solution, including features, integration, constraints and what's out-of-scope

- Core Features:
  - Apply a style to an image/video
  - Add new styles
- Integration:
  - Allow integration of images from the web or artists
- Alternatives:
  - Provide an alternative for users to manually add their own images
- Constraints:
  - Low latency: Maintain a low inference latency below 10 seconds for a given frame. We can accelerate the frame-by-frame processing of videos with batching.

    - Approved styles: Restrict the use of styles to a curated list, avoiding aggressive or offensive styles to be added to the platform.
    - Training duration of the model on a new style (+/- 2hours on a GeForce RTX 3080 Nvidia GPU)
- Out-of-scope:
    - Possibility to change style for each different frame.
    - Possibility to select a specific moment on which apply to the style, the user must only give the video / the part of the video he wants to style.
    - Improved support for any video format and resolution by using dedicated supplementary neural networks.

FEASIBILITY

Discuss the feasibility of the solution and if we have the required resources.

The project seems totally feasible as we have already acquired all the data and a first version of the model has been tested and has shown us to be able to do what we need it to do. Furthermore, we have access to the GPUs from ALAN cluster which contains everything we need in terms of computational power.

DATA

Identify the training and production data sources, as well as the labeling process and decisions.

Our models do *not* require labelled data, which simplifies data collection, processing and usage.

Training:

- A large dataset of arbitrary and diverse images (e.g., COCO, LAION-5B), that can be reused for each trained style.
- For each style, a single or several reference images.

Inference:

- A trained model can be applied to any image or video (i.e., frame by frame).

Validation:

- We will validate our models on video datasets (e.g., Kinetics-600)

METRICS

Prioritize key metrics that reflect the objectives

- ***Visual Quality***: This involves subjective evaluation by humans to determine how visually appealing the stylized videos are compared to the original input videos. Feedback can be gathered through user studies or subjective ratings.
- ***Fidelity to Style***: Stylized videos should be accurately modified with the reference style. We will use a perceptual metric that compares the style of two images by computing differences of internal feature maps of pre-trained large convolutional networks, like training.

- **Temporal Consistency**: Does the style remain consistent across frames in the video? Jitteriness or flickering can indicate poor temporal coherence. Metrics like temporal consistency and motion coherence can be used to measure this aspect.
- ***Computational Efficiency***: How fast can the model stylize videos? This is crucial for real-time or near-real-time applications. Metrics such as frames per second (FPS) or processing time per frame can be used to evaluate computational efficiency.
- **User Satisfaction**: Ultimately, user satisfaction is a crucial metric. Gathering feedback from users about their experience with stylized videos can provide valuable insights into the overall quality and usability of the model.

EVALUATION

Design offline and online evaluation criteria

- Offline: We will assess visually the quality of our network on a set of videos not involved in the training. We will also test the system latency by executing performance tests.
- Online: Users can provide feedback on their generated videos using the interface. This includes the overall quality of the video, how consistent the generated video is with respect to the style and through time, ...

MODELING

List the iterative approach to model our task

- Scrapping the data on the web
- Train the model to obtain the different styles
- Apply the model on the user content and make him happy!

INFERENCE

Decide whether we want to do batch (offline) or real-time (online) inference

We want to perform real-time inference where the videos are fed to the model by the users at any time to produce the stylized videos. This requires handling the request traffic.

FEEDBACK

Outline sources of feedback from our system to use for iteration.

/

PROJECT

Define the required team members, deliverables and projected timelines.

- *Team members:*
    - *Boveroux Laurie*
    - *Lewin Sacha*
    - *Louette Arthur*

    - *Mangeleer Victor*
    - *Schyns Axelle*
- Deliverable: **ClipMorph**
- *Project timelines: See project deadlines*