1. Title Research Project: Machine Learning in Video Editing
2. AMS Mathematical Subject Classification: 68T45 Machine vision and scene understanding
3. ACM Computing Reviews Categories and Subject Descriptors: I.2.10 Video analysis

**The Anatomy of Video Editing: A Dataset and Benchmark Suite for AI-Assisted Video Editing**

**Relevance**

I will use this paper to understand and how to work with the dataset, which will be used further in the training of the AI model. Also, it will help me understand better how to divide a video in shots and how to classify each shot in order to have a more accurate program.

**Structure**

Abstract

1. Introduction
2. Related Works
3. Anatomy of Video Editing: Dataset
   1. Shot Attributes
   2. Scene Composition and Camera Setups
   3. Annotation Procedure
   4. Dataset Statistics
4. Anatomy of Video Editing: Benchmark Suite
   1. Shot Attributes Classification
   2. Camera Setup Clustering
   3. Shot Sequence Ordering
   4. Next Shot Selection
   5. Missing Shot Attributes Prediction
5. Experimental Results and Discussion
   1. Experimental Results
6. Conclusion
7. References

**References**

Citatii: 0, Referinte: 50. Autorii fiecarei referinte sunt ordonati alfabetic, iar referintele sunt ordonate alfabetic dupa autori. Formatul referintelor este   
nr Referinta. Autori: Publicatie (An) Pagini in care s-a folosit

**STEP: Segmenting and Tracking Every Pixel**

**Relevance**

This paper is useful as the segmentation is one of the most important part of this research, because we want to edit subjects differently (e.g. we want to increase the blue saturation for the sky, but now for the water in a beach shoot), so in this matter we can approach each subject with its own settings. This will help me understand how to integrate the segmentation model.

**Structure**

Abstract

1. Introduction
2. Related Works
3. Datasets
4. Metric
   1. Metric Analysis
   2. STQ Metric
5. Baselines
6. Results
7. Conclusion
8. References

**References**

Citatii: 0, Referinte: 69. Ordinea referintelor este una aleatorie. Formatul referintelor este   
[nr Referinta] Autori. Publicatie. SursaPublicatie. An. Pagini in care s-a folosit

**Video Super-Resolution via Deep Draft-Ensemble Learning**

**Relevance**

This paper will provide a deeper understanding of how to increase the video quality for a better segmentation, so even the poorly shot videos can be renewed and look like a movie. Also, it will increase the knowledge on how to use CNNs in video editing and how much it can be improved.

**Structure**

Abstract

1. Introduction
2. Related Works
3. SR Draft-Ensemble and Its Analysis
   * 1. SR Draft Generation
     2. SR Draft-Ensesmble and Its Visualisation
     3. The Statistics of SR Draft Ensemble
4. Ensemble Learning via CNN
   * 1. Motivation
     2. Network Architecture
     3. CNN Training
5. Experiments
   * 1. Implementation
     2. Validation of Our Network
     3. More Analysis
     4. Running Time
6. Conclusion
7. Acknowledgements
8. References

**References**

Citatii: 0, Referinte: 21. Ordinea referintelor este una aleatorie. Formatul referintelor este   
[nr Referinta] Autori. Publicatie. SursaPublicatie, An.

**Dense Pixel-to-Pixel Harmonization via Continuous Image Representation**

**Relevance**

This will provide another way to improve the video quality, this method being excelent for action videos, where pixels can be lost due to lots of movements. This also provides datasets that can be used in the training of the model.

**Structure**

Abstract

1. Introduction
2. Related Works
3. Proposed Method
   1. Overview
   2. Decoupled Content and Appearance MLPs
   3. Low-Resolution Image Prior
   4. HR Image Harmonization
4. Experiments
   1. Experimental Settings
   2. Comparison with Existing Models
   3. Ablation Studies
5. Applications
6. Limitations
7. Conclusion
8. References

**References**

Citatii: 0, Referinte: 44. Ordinea referintelor este una aleatorie. Formatul referintelor este   
[nr Referinta] Autori. Publicatie. SursaPublicatie cu italic, Pagini folosite, An. Pagini in care s-a folosit

**Deep Interactive Object Selection**

**Relevance**

This will help me have a better understanding of how object selection works and how we can integrate it in a segmentation models, improving the overall performance of the model that will be used in the video editing

**Structure**

Abstract

1. Introduction
2. Related Works
3. The proposed algorithm
   1. Transforming user interactions
   2. Simulating user interactions
   3. Fine tuning FCN models
   4. Graph cut optimization
   5. Evaluation and complexity
4. Experiments
   * 1. Settings
     2. Results
     3. Comparisons tosemantic segmentation approaches
     4. Segmenting object parts
     5. Refinement by Graph Cut
5. Conclusion
6. References

**References**

Citatii: 0, Referinte: 26. Ordinea referintelor este una aleatorie. Formatul referintelor este   
[nr Referinta] Autori. Publicatie. SursaPublicatie cu italic, Pagini folosite, An. Pagini in care s-a folosit

**Structure Idea**

Abstract

1. Introduction
2. Related Works
3. Dataset
4. Proposed Method
5. Experiments
6. Conclusions
7. References

**Working Hypothesis**

Improving automated video editing by dividing each task to a distinct network, such as VSRN(improve resolution), VDN(removing noise), or VCCGAN(color grading), etc. and merging the results.

**Methodology**

1. Creating the dataset
   1. For the scene detection network, collect a large dataset of videos labeled with different scene types, such as indoor, outdoor, day, night, etc.
   2. For the object detection network, collect a large dataset of images labeled with different object types, such as people, cars, animals, etc.
   3. For the sound detection netowrk, collect a large dataset of audios labeled with different actions such as walking, opening a door, etc.
2. Feature Extraction
   1. Extracting the audio from the video to used it for the color grading.
   2. Diving the video in scenes.
3. Network Selection
   1. Video super-resolution (VSRN): If the input video is low resolution, use a VSRN to upscale the video to high resolution.
   2. Video denoising (VDN): If the input video is noisy, use a VDN to remove the noise.
   3. Video color correction (VCCGAN): If the input video has poor color quality, use a VCCGAN to correct the colors.
4. Enhancement – for each scene in the video
   1. Video super-resolution (VSRN): If the input video is low resolution, use a VSRN to upscale the video to high resolution.
   2. Video denoising (VDN): If the input video is noisy, use a VDN to remove the noise.
   3. Video color correction (VCCGAN): If the input video has poor color quality, use a VCCGAN to correct the colors.
5. Output video generation
   1. Combine the enhanced scenes to generate the output video.
6. Evaluation
   1. Evaluate the quality of the output video using a variety of metrics, such as perceptual similarity index (PSI), structural similarity index (SSIM), and video quality metric (VQM).

**Original Approach**

The use of GAN for the integration of audio in video editing based on each objects that appear in a scene.

**Experiments Description**

* Parallelization of the training for each network.
* Evaluate the performance of individual networks.
* Evaluate the performance of the proposed approach.
* Combining the enhanced scenes.

**Possible Original Contribution**

How to optimise real-time video editing? This method can be applied to a range of video applications, including surveillance, providing sharper images for security to monitor, and filmmaking, reducing the time spent on manual video editing.

**References**

[1] Dawit Mureja Argaw, Fabian Caba Heilbron, Joon-Young Lee, Markus Woodson and In So Kweon - The Anatomy of Video Editing: A Dataset and Benchmark Suite for AI-Assisted Video Editing. <https://www.ecva.net/papers/eccv_2022/papers_ECCV/papers/136680195.pdf>

[2] Mark Weber, Jun Xie, Maxwell Collins, Yukun Zhu, Paul Voigtlaender, Hartwih Adam, Bradley Green, Andreas Geiger, Bastian Leibe, Daniel Cremers, Aljosa Osep, Laura Leal-Taixe, Liang-Chieh Chen - STEP: Segmenting and Tracking Every Pixel. <https://arxiv.org/pdf/2102.11859.pdf>

[3] Renjie Liao, Xin Taom Ruiyu Li, Ziyang Ma, Jiaya Jia - Video Super-Resolution via Deep Draft-Ensemble Learning. <https://openaccess.thecvf.com/content_iccv_2015/papers/Liao_Video_Super-Resolution_via_ICCV_2015_paper.pdf>

[4] Jianqi Chen, Yilan Zhang, Zhengxia, Keyan Chen, Zhenwei Shi, Beihang University- Dense Pixel-to-Pixel Harmonization via Continuous Image Representation. <https://arxiv.org/pdf/2303.01681.pdf>

[5] Ning Xu, Brian Price, Scott Cohen, Jimei Yang, Thomas Huang - Deep Interactive Object Selection. <https://arxiv.org/pdf/1603.04042.pdf\\>

[6] John L. Gibbs - Video Color Grading Via Deep Neural Networks. <https://www.iadisportal.org/ijcsis/papers/2018130201.pdf>

[7] Jiandan Zhong, Tao Lei, Guangle Yao - Robust Vehicle Detection in Aerial Images Based on Cascaded Convolutional Neural Networks. <https://www.researchgate.net/publication/321280709_Robust_Vehicle_Detection_in_Aerial_Images_Based_on_Cascaded_Convolutional_Neural_Networks#pf9>

[8] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang - Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. <https://arxiv.org/pdf/1608.03981.pdf>

[9] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta and Anil A Bharath - Generative Adversarial Networks: An Overview. <https://arxiv.org/pdf/1710.07035.pdf>

[10] Jie Gui, Zhenan Sun, Yonggang Wen, Dacheng Tao, Jieping Ye - A Review on Generative Adversarial Networks:

Algorithms, Theory, and Applications - <https://arxiv.org/pdf/2001.06937.pdf>

[11] Kieran G. Larkin – Structural Similarity Index SSIMplified: Is there really a simpler concept at the heart of image quality measurement? <https://arxiv.org/ftp/arxiv/papers/1503/1503.06680.pdf>

[12] Alice Lucas, Santiago Lopez-Tapia, Rafael Molina and Aggelos K. Katsaggelos - Generative Adversarial Networks and Perceptual

Losses for Video Super-Resolution. <https://arxiv.org/ftp/arxiv/papers/1806/1806.05764.pdf>

[13] Ming-Jun Chen, Alan C. Bovik - Fast structural similarity index algorithm.

[14] Panagiotis Kouzouglidis, Giorgos Sfikas, Christophoros Nikou – Automatic Video Colorization Using 3D Conditional Generative Adversarial Networks. <https://arxiv.org/pdf/1905.03023.pdf>

[15] Hongying Liu, Zhubo Ruan, Peng Zhao, Chao Dong, Fanhua Shang, Yuanyuan Liu, Linlin Yang, Radu Timofte - Video Super-Resolution Based on Deep Learning: A Comprehensive Survey <https://arxiv.org/pdf/2007.12928.pdf>