Journal of Cultural Heritage 68 (2024) 298–306

Contents lists available at ScienceDirect

Journal of Cultural Heritage

journal homepage: www.elsevier.com/locate/culher

Original article

An intelligent epistemological tool for audiovisual analysis and mediation of video art archive

Jana Horakova[a], Pavel Sikora[b,*], Stepan Miklanek[b], Dusan Barok[c], Jiri Schimmel[b], Kamil Riha[b]

[a] Masaryk University, Faculty of Arts, Dept. of Musicology, Arna Novaka 1/1, Brno, 60200, Czech Republic
[b] Brno University of Technology, Faculty of Electrical Engineering and Communications, Dept. of Telecommunications, Technicka 12, Brno, 616 00, Czech Republic
[c] Independent researcher, Norway

ARTICLE INFO

Article history:
Received 25 September 2023
Revised 12 May 2024
Accepted 20 June 2024

Keywords:
Analysis
Artificial intelligence
Augmented iconography
Deep learning
Intelligent epistemological tool
Video art

ABSTRACT

This article focuses on the development and application of intelligent software for image and sound recognition to perform iconographic and audiographic analyses of the work of video art pioneers Steina and Woody Vasulka (the Vasulkas). The AI epistemological tool Vasulka Live Archive is designed to provide unique results that benefit from the synthesis of automatic statistical analysis across the dataset and application of predefined categories that are the results of aesthetic evaluation of the Vasulkas' videos and inspired by terminology of video art aesthetics (Weibel, Krauss). The advantages of this AI tool reveal themselves particularly when the tool is used for transmedia analysis of the whole dataset (the Vasulkas' work) as the accuracy and completeness of its results are out of reach of an individual human researcher. We argue that this kind of AI tools can contribute to more exact and data-based findings on media art aesthetics, it can contribute to establishing a new field augmented iconology (Spratt) as well as expanding the sphere of AI tools application towards digital collections of experimental and conceptual art of 20th and 21st centuries.

1. Introduction and research aim

Our efforts are in response to the ever evolving museum and archival landscape, as well as the general tendency toward digitizing cultural artifacts (see Google Art & Culture[1], 2011; Europeana[2], 2009). Open access to the digitalized and interconnected art collections of notable institutions (e.g. Tate Britain; Mo Ma, The Museum of Modern Art; Rijk Museum) has recently increased scholars' interest in utilizing machine learning to handle these massive datasets.

1.1. Related works

An increasing number of projects focus on the development of intelligent tools for the automated examination of digitalized art collections [1–7]. Multiple interdisciplinary research teams have adapted, created, and combined a range of Artificial Intelligence (AI) technologies to serve as a tool for the study of art history and cultural heritage.

The focus has been on older time periods and collections of traditional artifacts such as paintings, drawings, or sculptures [1–5]. Part of these experiments were conducted on collections of artistic photographs [8]. However, as far as we know, we present the first application of machine learning on a collection of video art.

These projects typically focus on different kinds of sorting and searching of static images in databases of visual art, and they are applied to datasets of artworks from older historical periods, with their realistic symbolic vocabulary based on mimetic depiction of objects and environments (regardless of whether the subject matter of the artwork is taken from the observation of the world or is of allegoric, religious, or mythological nature) [9,10]. The results of these projects serve as evidence that AI tools are able to automatically accomplish goals specified by human art historians almost a century ago (such as Wolfflin's methods of style distinction and analysis [9,10], and Panofsky's iconography as used by Spratt [8]).

Due to the conceptual nature and formal diversity of modern art, the applicability of machine learning to the art of the 20th and

* Corresponding author.
E-mail address: pavel.sikora@vut.cz (P. Sikora).
[1] artsandculture.google.com.
[2] europeana.eu.

https://doi.org/10.1016/j.culher.2024.06.011

21st centuries is regarded challenging. This notion is confirmed by the unsatisfactory results of trials involving the application of AI algorithms to current-art databases [8]. However, we are convinced that the issue is less related to the nature of contemporary art and more to the fact that AI software was trained on datasets, such as ImageNet [11], which are incompatible with the visual language of most modern and contemporary art [8].

1.2. Main goals of our research

In response to the current situation described above in the field of AI application to digitalized art archives, we decided to develop AI software specially trained to be an effective analytical tool when applied to audiovisual artworks (artistic videos) with very specific aesthetics compared to the pop-culture imagery contained in ImageNet. Our research aims to train AI software to become an expert in the iconography of Woody and Steina Vasulka's experimental creative videos.

The specially designed web portal Vasulka Live Archive [12] is the main result of the research and is available online for the art historian community to serve as a unique intelligent epistemological tool for AI-assisted iconological analysis of the Vasulkas artistic videos. The online, open-access, intelligent epistemological tool can (a) aid in the intertextual structural analysis of the videos (relations between visual and acoustic leitmotifs, usage of various tools and software, or media); (b) quantify the degree of ambiguity (multivocal and metaphorical imagery) in these videos; and (c) assist in determining the amount to which the video demonstrates a tendency toward abstraction.

The rest of the article is structured as follows. Section 2 describes the dataset and the AI-based methods for the analysis of the audiovisual content of the Vasulkas' videos. Next, we summarize the main results in Section 3. Finally, we provide a discussion and a conclusion in Sections 4 and 5, respectively.

2. Materials and methods

2.1. Dataset characteristics

We worked with videos by Steina and Woody Vasulka, stored in a digital archive in possession of the Center for New Media Arts—Vasulka Kitchen Brno.[3] The total data volume is 536 GB, including 1,800 audiovisual works. After discarding redundant content, we ended up with 880 videos by Vasulkas, a 137 GB corpus of material that would take nearly seven days to watch.

Since there is no generally recognized list of the Vasulkas' works that we could use as a reference when identifying videos in the dataset, we had to create our own. We accomplished this by comparing and overlapping the lists of Vasulkas' works available on the web sites of the authoritative institutions.[4] In this way, we compiled a list of 108 video titles, of which 21 were made by Woody Vasulka, 49 by Steina Vasulka, and 38 works were made by both authors. We were able to identify 105 distinct works by the Vasulkas, represented by 124 videos and video-documentaries (62 by Steina Vasulka, 10 by Woody Vasulka, 38 by Woody and Steina Vasulka, and 14 made in cooperation of the Vasulkas with other authors). We can therefore assert that the collection of their works we worked with is nearly complete. Moreover, the vast majority of videos are available as complete versions (master tapes), with only a few appearing as fragments in the dataset.

In terms of **visual information**, the Vasulkas' work is varied, comprising a small volume of documentary footage and the majority of manipulated images. While the Vasulkas' earliest videotapes made use of only camera effects, later works showed a tendency toward blending documentary footage with visual effects; these were created by disrupting electronic signal using customized synthesizer-based instruments and a range of hardware and software tools.

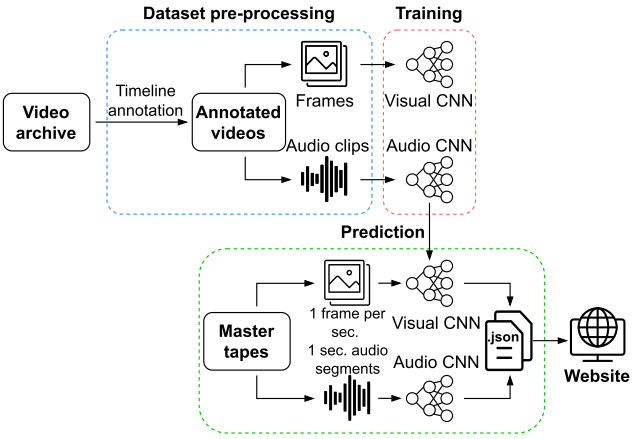


Fig. 1. Flowchart of our system for audiovisual content recognition.

The **sonic character** of Vasulkas' work is predominantly made up of various kinds of noise and synthetic sounds that can be classified as electronic music. Voice (talking, singing) and its manipulation in real time contribute significantly to the soundscape of their videos. Another typical sound is produced by Steina Vasulka playing a violin, later on a Musical Instrument Digital Interface (MIDI) violin, which Steina used to control image sequences in real time. There are also sounds of natural elements, such as wind, waterfalls, or fire, as well as some industrial noises, such as the sound of a car engine.

The extensive manipulation of images and sounds in Vasulkas' work is an expression of the tendency towards abstraction and the detachment from the recorded reality in favor of synthetic, artificial images and sounds [13,14]. For the goals of dataset sorting and training the Neural Networks (NNs) to perform an iconographic and audiographic analysis of videos automatically, these categories were required to be separated.

During dataset creation, a sub-dataset was created for experimental verification of classification using NNs. The dataset contains a small number of visual classes with a limited number of frames. The results of these experiments were published in [15].

The dataset with which we worked could not be published in its entirety for licensing reasons, but the videos we identified as master tapes of the Vasulkas work are fully available on the Vasulka Live Archive. The two software trained on this dataset described in the following section are available in the open access regime on GitHub.[5,6]

2.2. Audiovisual content recognition

We used Convolutional Neural Networks (CNNs) to develop our audiovisual content recognition system because they are the most common choice for image and audio classification [16]. Before deploying the system (see Fig. 1), we had to correctly prepare the training dataset for CNNs. To organize and categorize the dataset of the Vasulkas' recordings, we developed a website-based interface that allowed for automatic searching across the full unsorted video

[3] vasulkakitchen.org.

[4] vasulka.org, Foundation Langlois, LI-MA, Electronic Art Intermix, Berg Contemporary, monoskop.org.

[5] github.com/vasulkalivearchive/audio.

[6] github.com/vasulkalivearchive/video.

archive. The interface made it simple to locate the master tapes or the most representative segments of the Vasulkas' artistic videos. The selected videos were provided with basic metadata (author, work title, year, abstract, etc.). In addition, it was made possible to simultaneously play these videos and annotate the audio and visual categories. Annotations were produced with a second precision on the timelines, and as the majority of the videos featured many categories simultaneously, it was possible to use numerous timelines for this purpose. Before training the CNNs, the audio and image datasets were separated into training and validation sets.

Subsequently, two CNNs were trained—one for visual categories and one for audio categories. First, a CNN was trained on the audio clips from the videos. In the case of the visual CNN, a multilabel image classifier was trained. Predictions were made on the timelines of all Vasulkas master recordings. Some parts of the master recordings are already included in the CNN training dataset. The videos for the training and validation subset were specifically chosen so that all categories occur in both subsets. Subsequent predictions were made on all master tapes.

The predictions of both trained CNNs were then uploaded to the Vasulka Live Archive website, where they are stored in the database for quick access. For our purposes, the high temporal accuracy of the classification was not crucial. Therefore, we chose to predict the occurrence of objects with one-second skips. Furthermore, predictions were used for a feature of the website that automatically suggests similar videos as described in Section 3.4.1.

2.2.1. Visual classes

Our dataset has twenty **visual classes** (see Table 1), some of which cannot be found in general-purpose datasets commonly used for CNN learning. Moreover, their distribution is quite unbalanced (see Fig. 2b).

For example, the class *Interior* was detected more than ten thousand times in the images of the dataset. In contrast, the frequency of the class *Car* does not exceed 500. The imbalance of class appearances in the dataset was compensated for by using weights. Before CNNs were trained, the number of images per class was determined and the weights were modified accordingly.

Moreover, the visual classes appear as individual objects, or there are several of them in one frame in the dataset. An example of the appearance of several classes at the same time in the frame shows Fig. 3a.

Table 1
Percentual F_1 score on a validation subset for the trained visual models.

Class name	Xception	NASNet	Inception
Body	84.5	88.6	88.7
Digit	88.0	**99.9**	**99.9**
Effect	49.7	39.4	39.8
Air	68.6	72.2	72.5
Earth	59.6	58.8	55.6
Fire	56.3	39.0	69.9
Water	71.6	70.6	66.0
Face	**90.1**	90.5	89.2
Keying	62.6	59.9	53.2
Interior	81.9	79.7	79.8
Landscape	81.5	80.4	74.8
Letter	87.7	89.8	84.0
TV set	78.9	78.5	85.6
Car	74.4	92.0	89.8
Fisheye	85.0	87.0	82.0
Rutt/Etra	61.4	71.0	68.1
Steina	34.9	63.6	28.6
Stripes	66.0	66.0	80.4
Violin	50.2	55.5	55.2
Woody	82.5	91.4	96.1
Macro average	70.8	**73.7**	73.0

Table 2
Percentual F_1 score on a validation subset for the trained audio models.

Class name	CNN14 (3 s)	CNN14 (1 s)	CNN14 (0.5 s)
Music (acoustic)	84.9	86.0	86.9
Music (electronic)	89.2	91.2	90.2
Music (violin)	83.3	84.0	81.9
Music (vocal)	83.6	79.6	78.1
Noise	92.6	93.4	93.1
Noise (air)	81.0	78.6	78.9
Noise (car)	83.9	77.0	83.1
Noise (fire)	90.9	**96.6**	**97.9**
Noise (water)	86.4	86.8	85.6
Speech	**94.0**	93.4	84.7
Macro average	**87.0**	86.6	86.0

It is the screenshot from the video *Summer Salt* [17], in which categories *Air, Body, Earth, Face, Machine vision, Lanscape*, and *Steina* were identified. On the other hand, Fig. 3b shows a case of a single class in the frame. It is the *Earth* class identified in the video *Lava and Moss* [18].

2.2.2. Audio classes

The **audio classes**, which have been used for the CNN training (see Table 2), consist of ten individual categories. The dataset is also imbalanced (see Fig. 2a). For example, the most frequently identified class is *Music (electronic*), that was recognized on more than three hours of videos. On the other hand, the occurrence of the class *Noise (fire)* was identified the least frequently, on a total of three and a half minutes of video length.

The next step was to standardize the length of the audio segments for training because manual annotation and subsequent extraction produced audio files of varying lengths. To do so, the audio dataset was divided into training segments of 0.5, 1, and 3 s in length.

2.3. Towards two complementary NN models

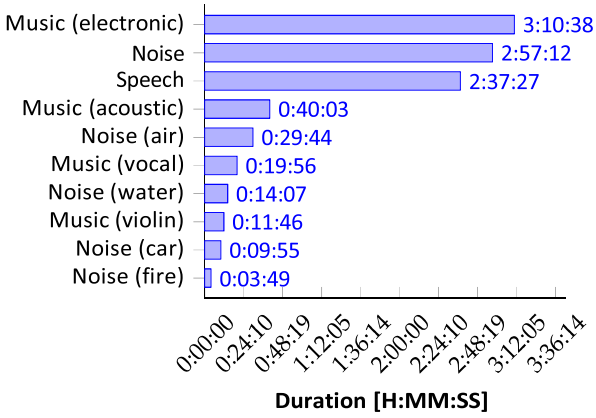
The categories for the training of the two NNs were determined considering the intended next step, which was to incorporate them into the Vasulka Live Archive website's interface to serve as a tool for providing users of the tool with an interface that enables simultaneous analysis of audio and visual motifs of the videos as they are audiovisual artworks.

2.3.1. The visual CNN

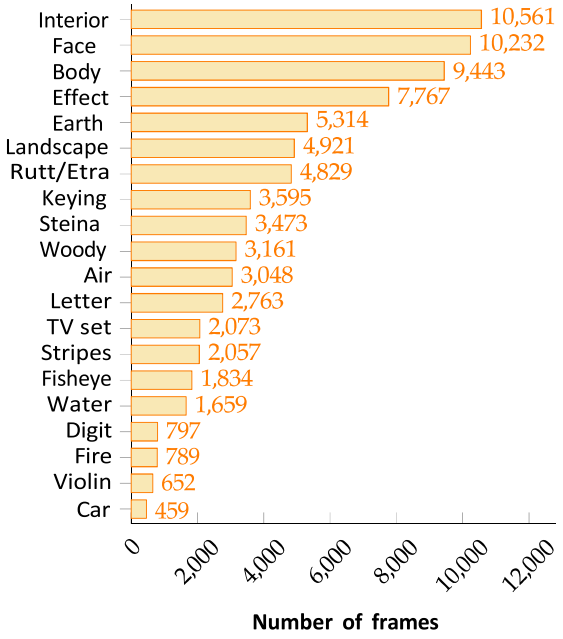
We decided to test three CNNs: Xception [19], NASNetLarge (NASNet-A) [20], and Inception-ResNet-v2 [21]. The core of Xception is made of depth-wise separable convolutions. The model is based on several improvements of the Inception model, on which the name Extreme Inception is based. This model is based on the hypothesis that it is appropriate to decouple spatial and channel convolutions on channels.

Then, NASNetLarge was tested. It is the best performing network trained on CIFAR [22] and ImageNet [11] datasets. The framework allows for the automated composition of individual blocks of the whole NN, which enables it to achieve the best results on a given dataset.

Finally, the Inception-ResNet-v2 model was applied experimentally to the dataset. We chose this model because the network achieves very good results on the ImageNet database and thus ranks among the state-of-the-art classifiers. Moreover, the model achieves acceptable computational requirements. The authors claim that residual connections dramatically improved training speed but did not increase accuracy. Thus, the accuracy of the model is achieved by a special design of the individual blocks of



(a) Occurrence of individual audio classes in the dataset.



(b) Occurrence of individual visual classes in the dataset.

Fig. 2. Dataset statistics.

(a) *Summer Salt* [17].

(b) *Lava and Moss* [18].

Fig. 3. Images from two videos.

the architecture based on the previous versions of the Inception architecture [21].

When we developed these three CNNs, we progressed as follows. We removed the last classification layer from all CNNs and replaced it with a *2D global average pooling* layer followed by a *dropout* layer. Next, we used a *fully connected* layer with a neuron activated by a sigmoid for each class. All models were trained for 15 epochs. For the predictions presented on the Vasulka Live Archive website, we chose the NASNetLarge model, as it had the best classification accuracy.

2.3.2. The audio CNN

In recent years, large-scale audio datasets such as AudioSet were classified using NNs [23]. In addition to CNNs [24], architectures such as convolutional recurrent neural networks (CRNNs) [25], or NNs with attention mechanisms were used in the task of audio classification [26]. We decided to use the CNN14 model presented in [27] due to its performance on large datasets and in transfer learning settings.

The CNN14 model comprises six convolution blocks and a fully connected layer with a size of 527 neurons, which corresponds to the number of classes in the AudioSet. However, the CNN14 model had to be customized to fit the parameters of our dataset. The number of neurons of the last layer was changed according to the number of audio classes we used for our dataset, which was ten (see Table 2). The CNN14 model contains a logarithmic mel spectrogram layer [30] on the input, which transforms audio into the frequency domain. We apply 64 mel filters to calculate the log mel spectrogram. Each convolution block consists of two layers activated by a ReLU function [31], then followed by an average pooling and a dropout of 0.2.

The output of the last convolution block is reduced to a one-dimensional embedding by a combination of max and mean pooling followed by a dropout of 0.5. The embedding is then activated by a ReLU function and follows a final dropout of 0.5. The predic-

tions are made by a fully connected layer with 10 neurons, which is activated by a log softmax function [32]. We trained the model three times for 60 epochs, each time with a different length of training segments. The negative logarithmic likelihood loss criterion [33] was used to train the model.

3. Results

3.1. Visual CNNs

Table 1 summarizes the results of all three CNNs tested for visual class analysis. The best model in terms of the macro-average F_1 score was the NASNetLarge model with 73.7%. The worst results were achieved by the Xception model with a margin of 2.9%.

To illustrate the results summarized in Table 1, two examples are used: Fig. 4a shows a screenshot from the video *Bent Scans* [29], in which Steina Vaulka performs different movements in front of the projector. The body is distorted and bent using Image/ine software (Tom Demeyer) in real time, and the image is instantly projected onto a wall. The CNN correctly identified classes *Interior* (100 %), *Body* (100 %), and class *Effect* (100 %), which means that some image manipulation took place in the frame. Fig. 4b shows a frame from the video *1-2-3-4* [28]. The video depicts animated numerical birthday cake candles 1, 2, 3, 4 on the striped background. The numbers and stripes change their positions rhythmically and overlap with each other. To do so, the author used Multi-Level Keyer, an analog tool with digital control, built by Georg Brown. The CNN successfully identified the class *Stripes* (100 %), and the class *Digit* (100 %) in the image.

3.2. Audio CNNs

The CNN model for audio content recognition was trained on sound segments of varying length. The results indicate that the quality of the classification is comparable across all segment lengths. Some classes scored a much poorer classification accuracy, which was primarily attributable to their underrepresentation. Several classes also contained audio material that was difficult for humans to distinguish, so even after manual correction, some classes contained material that was incomprehensible for NNs. Despite these shortcomings, a balanced level of categorization accuracy was achieved across all chosen sound classes. Table 2 displays the results of the predictions on a validation subset.

Due to the sufficient temporal accuracy of audio track tagging, we chose the CNN14 model, which was trained on 1-second segments, for prediction purposes on the master tapes of the Vasulkas.

3.3. CNNs in a service of an audiovisual analyses of the Vasulkas' videos

The CNNs we have trained are unique intelligent tools that enable investigators to perform automatic content-oriented (leitmotifs) and form-oriented (special effects usage, image-sound interactions, etc.) analysis of the artistic videos, both individually and across the dataset.

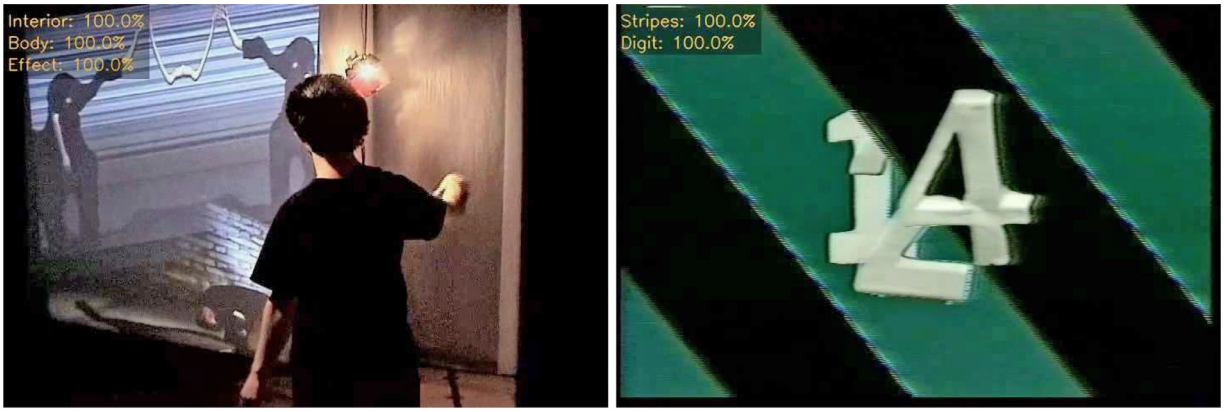
To increase the effectiveness of the CNN analysis, the models are integrated into a single interface of the Vasulka Live Archive website so that users can investigate the visual and audio aspects of the videos simultaneously. The simultaneous usage of the CNNs enables users to conduct the structural analyses of the videos, e.g. to observe how the representation of objects by visual and audio features is mutually supportive (it is usually the case of documentary videos), or how the visual and audio objects mutually convey a dominant position within the audiovisual works. An example of the case where the audio contains the mimetic information while the visual information is highly abstracted is shown in Fig. 5b. It is an image from the video *Telc* [35]. Significant usage of the Rutt/Etra synthesizer almost obliterates the visual information, while the audio carries the information with a documentary-like quality, allowing us to discern that someone is remarking on a car driving along a street. Fig. 5a is a still from the video *Violin Power* [34] that illustrates the opposite situation. The image depicts Steina Vasulka playing a violin, but the audio clarifies that it is a MIDI violin, i.e. an electronic device used as an interface for manipulating electronic signals.

3.4. Intelligent epistemological tool for audiovisual analysis of Vasulkas' videos

The specially designed Vasulka Live Archive website was created to harness the full potential of specially trained CNNs for audio and visual analysis of Vasulkas' videos. This applies particularly to the *Machine Learning* section, which we refer to as an interactive epistemological tool, while the *Machine Vision* section serves educational purposes, providing a basic insight into the hidden processes of AI software.

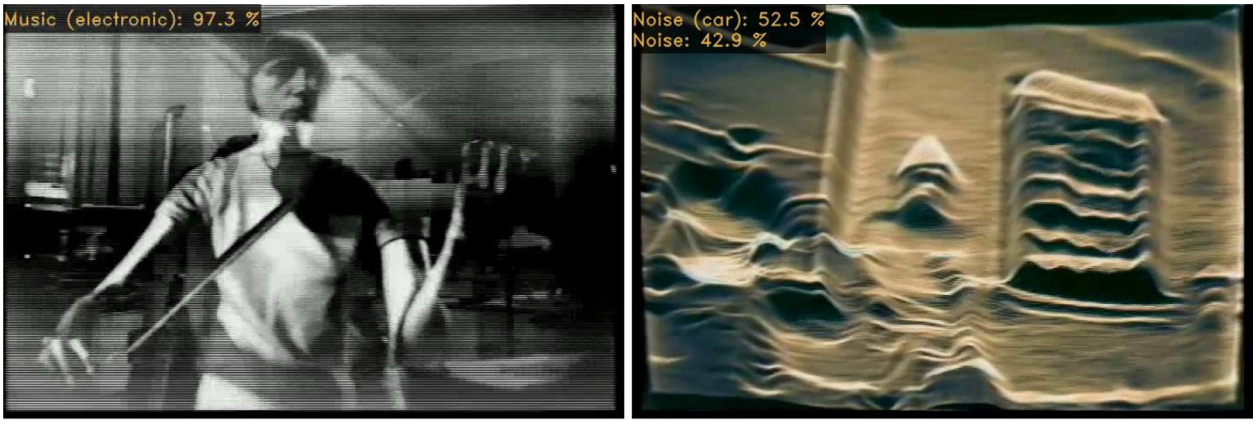
3.4.1. Machine learning

In the Machine Learning section of the website, users can filter videos according to the appearance of visual and audio classes in the videos, with frequencies calculated as a percentage of the appearance of a given class with a probability threshold of 1%. For example, in the video *1-2-3-4* [28], the *Digit* class was identified in 95% of the frames and therefore the weight of this class is 0.95.



(a) Interior: 100 %, Body: 100 %, Effect: 100 %. (b) Digit: 99.9 %, Stripes: 99.7 %.

Fig. 4. Image from 1-2-3-4 [28] (on left), Image from *Bent Scans* [29] (on right).



(a) Music (electronic): 97.3 %. (b) Noise (car): 52.5 %, Noise: 42.9 %.

Fig. 5. Image from Violin Power [34] (on left) and Telc [35] (on right).

Fig. 6. An example of the dominance of the acoustic over the visual in video Swan Lake [38] (Violin: 0 %, Music (violin): 100 %).

The class weight can be adjusted for several occurrence rates: 1 to 100%, 25 to 100%, and 50 to 100%. However, we expect that the ability to search for clusters of videos with shared features from 1 sec to whole video can bring main benefits. In this way, we can analyze the Vasulkas' work from the perspective of repeating visual and sound motifs and compare them in terms of their rarity or density of their occurrence.

The interface enables users to study the results of the video classification frame by frame in real time (see Fig. 6). This is made possible by the inclusion of a dashboard to the video player, which displays a list of visual and audio classes, including data on their frequency of occurrence based on weight (with per-video precision) and probability measure (with per-frame precision). During video playback, the cursor automatically advances along the horizontal timelines of each class.

In addition, an algorithm has been implemented to identify similar videos, calculated as the distance between the videos, each represented by a weight vector containing values corresponding to all visual and audio classes. A value of one signifies that a given class occurred for the whole duration of the video, and a value of zero represents no occurrence of a particular class. Therefore, a set of similar videos to a target video in terms of the visual and audio class occurrence is found by calculating the Manhattan distance between respective weight vectors. This metric was chosen to rank similar videos because it is inexpensive to compute and suitable for content-based retrieval, as shown in [36].

3.4.2. Machine vision

The Machine Vision section shows how the NNs can see the Vasulkas videos. It was inspired by the Activation Atlas project [37], which uses visualization to illustrate the inner workings of NNs of different layers.

We used the intermediate outputs from five of the 132 CNN layers from Xception model, which is enough to sufficiently illustrate the process of decomposition of the image by CNNs, that enables them to capture features of the image, to identify individual objects in it, and finally to predict their appearance in the given image. The website section *Machine Vision* was created using the JavaScript library Three.js [39], which allows the images to be mapped in the shape of a sphere (see Fig. 7a).

When the mouse hovers over any image mapped on the sphere, a list of objects recognized by the AI software is shown first. By clicking on the image, the CNN layers spread over the screen, making the process of gradual decomposition and analysis of the image perceivable, and it is accompanied by a list of categories, including a percentage of the prediction probability of a particular label. This feature is illustrated in Fig. 7b. It shows an image from the *Techne*

(a) 3D Sphere as a user interface.

(b) Expanded NN of the selected image.

Fig. 7. Illustrations from website.

and Eros video [40] and the four stages of its decomposition by the AI software.

The interface was inspired by a reflective sphere used in Vasulkas' installations. It is often placed in the center of optical apparatuses called *Machine Vision* [41] or *Allvision* [42], and works in cooperation with cameras recording the images of the surroundings, which are reflected in its mirrored surface. Similarly, the Vasulkas employed the fisheye effect's characteristically convex form refers to the human ability to enlarge its field of view through a machine's eyes (see, for example, *Summer Salt* [17]).

Steina Vasulka wrote about this:

> "In the late seventies, I began a series of environments titled Machine Vision and Allvision, with a mirrored sphere....These automatic motions simulate all possible camera movements that free the human eye from being the central point of the universe. [41]"

4. Discussion

The Vasulka Live Archive website not only provides access to the outputs of our AI driven research on the Vasulkas work. Our main goal was to create a unique intelligent tool that integrates the knowledge of iconography of the Vasulkas videos with the attributes of video art as defined by prominent theorists [14,43]. We can compare the website to a microscope or telescope, as it is a tool that allows us to see the Vasulkas work both in detail and in distance, which might bring new findings and knowledge about their work. This epistemological tool can be used by scholars, students, curators, or artists interested in this work.

4.1. Epistemological tool for audiovisual analysis of video art

The AI tool integrates two expert software, one for analyzing visual motifs and the other for analyzing audio motifs on the database of master tapes representing the Vasulkas' work. The tool can bring high effectivity and exactness into research focusing on an intertextual structural analysis of the videos (relations between visual and acoustic leitmotifs).

In general, a thorough structural analysis of any audiovisual work, such as video art, must involve an analysis of visual and aural components. It was found that the strategy is effective, particularly when the audiovisual material is substantially modified to gain the effect of abstraction (distraction from reality), as is the case of the Vasulkas videos. In these circumstances, the complementary utilization of the AI software in the web-epistemological tool led to more satisfactory findings, because their outcomes supported each other in a process of identifying (audio or visual) things.

4.2. Quantifying ambiguity

By using a special interface that allows one to simultaneously track the probability of occurrence of each class in the videos as it evolves over time, we can measure the degree of ambiguity with which the Vasulkas deliberately worked. The word ambiguity is used here to signify a certain way the Vasulkas implement multivocal and metaphorical imagery into their works using different kinds of montages of images (superpositions, keying, spatial and temporal montages). For example, it is not uncommon for an image to denote and indicate two or three categories with high probability at once, one of them being the *Effect* category, as we observed in our research.

For example, in the *Lilith* video [44], see Fig. 8, the NNs recognized the *Face* and the *Landscape* with almost 100 % probability, and the Effect category was also represented by almost 100 %. By comparing this result with the screenshot shown in Fig. 8, it shows that the face merges with the landscape, one overlaps the other, and vice versa, creating a metaphorical image of a face-landscape melting in the waves of an electronic signal, as if emerging from the author's memories.

Although the numerical representation of the composition of the image may seem imprecise, it is an accurate expression of what the viewer is actually seeing. The results of the analysis can be used for the quantification of ambiguity as a characteristic feature of the work of the Vasulkas.

4.3. Scaling abstraction

As the Vasulkas' videos oscillate between documentary depiction of reality by camera and its diverse manipulations by interferences into electronic signal, the CNNs not only be trained to recognize the visual or audio objects, but also to measure the degree of abstraction, frame by frame. To do so, users of the website can filter categories such as *Effect, Noise (audio)*, or *Rutt/Etra*. In addition, from the standpoint of the aesthetics of video art, it is frequently less important to recognize the objects on the video than to be

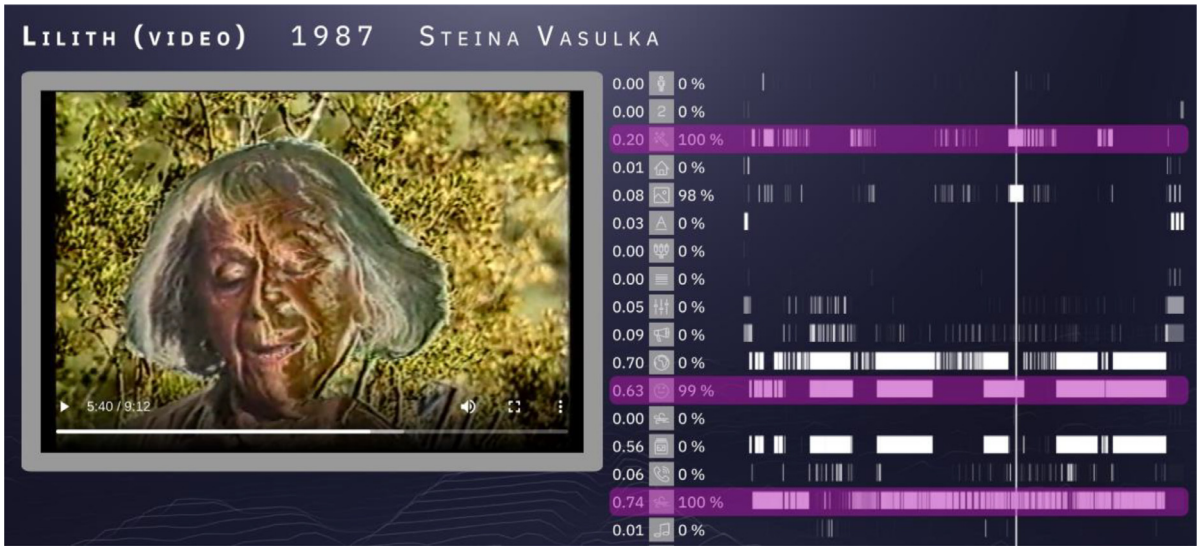


Fig. 8. Example of ambiguity [44], (Effect, Face and Landscape almost 100 %).

able to quantify the degree of transformation or level of abstraction of the image and/or sound. These measures can significantly contribute to accurate, data-driven aesthetic evaluations of electronic art and video art in particular.

5. Conclusions

The Vasulka Live Archive (comprising two AI software and the website) represents a successful achievement of the goal, which was to develop an intelligent epistemological tool (trained to recognize the recurring motifs and the main aesthetic categories of video art) for augmented iconology [8] of the Vasulkas life work.

The purpose was to investigate the potential of AI for the development of novel research methods and tools that will make it possible to implement augmented iconography, which can bring new and data-driven knowledge and interpretations of videoart.

The goals of our research are informed by Vasulkas' lifelong effort to investigate the nature of the electronic medium by means of artistic experiments (see video *Artifacts* [45]), and to reflect the broader social context within which electronic media and technology are embedded (see, for example, *The Art of Memory* [46]; or *The Brotherhood*, the interactive constructions, [47]). Vasulka Live Archive is the result of translation of their artistic exploration of video technology (as a medium of seeing and knowing) into the field of AI and machine learning.

We demonstrate that AI tools, when correctly trained for certain tasks, are not only useful for the automated search and sorting of datasets, but may also provide new knowledge of the studied artworks. We proved that the AI epistemological tool can help with (a) intertextual structural analysis of the videos, (b) measuring the degree of ambiguity (multivocal and metaphorical imagery) in these videos, and (c) determining how much a particular video exhibits a tendency toward abstraction. Its use is particularly effective when the research focuses on evaluating the work of the Vasulkas as a whole, or groups of videos, for which the results of working with the tool may be unattainable in other ways.

These three examples are not a complete list of how the tool can be used. In further research, we encourage researchers to experiment with the tool to test its potential, which has not been discovered so far. For researchers focusing on the application of AI in art and culture, it may also be interesting to observe the visualization of the work of intelligent software in real time, that is also incorporated feature of the website, and thus to study the machine vision in the process of perception. The Vasulka Live Archive contributes to the emerging field of augmented iconology that is based on the use of AI for iconographic analysis.

Original AI tool development contributes to the establishment of new research methods based on the tandem collaboration of the researcher with the intelligent tool and design of the interface that allows to observe the NN in the process of seeing the particular video (a kind of performance of machine vision); these indicate that our research is not only aimed at improving the historiographic knowledge of the Vasulkas work, but that it was also inspired by poetics of their work, particularly their approach to technology as a partner in a creative dialogue.

Acknowledgements

The article summarizes the main results of the Media Art Live Archive project: Intelligent Interface for Interactive Cultural Heritage Mediation (MediaArtLiveArchive, TL02000270) cofinanced with the state support of the TACR as part of the ETA program. The authors thank Vasulka Kitchen Brno and the Brno House of Arts for sharing the Vasulka digital archive and their trust. Special thanks to Masaryk University's Theory of Interactive Media students for their assistance in video tagging.

References

[1] N. Garcia, B. Renoust, Y. Nakashima, Context-aware embeddings for automatic art analysis, in: Proceedings of the 2019 on International Conference on Multimedia Retrieval, 2019, pp. 25–33, doi:10.1080/01621459.1958.10501452.
[2] G. Demir, A. Çekmiş, V.B. Yeşilkaynak, G. Unal, Detecting visual design principles in art and architecture through deep convolutional neural networks, Autom. Constr. 130 (2021) 103826, doi:10.1016/j.autcon.2021.103826.
[3] R. West, A. Burbano, et al., AI, arts & design: questioning learning machines, Artnodes (26) (2020) 1–9.
[4] C.B. El Vaigh, N. Garcia, B. Renoust, C. Chu, Y. Nakashima, H. Nagahara, GCN-Boost: artwork classification by label propagation through a knowledge graph, in: Proceedings of the 2021 International Conference on Multimedia Retrieval, 2021, pp. 92–100, doi:10.1145/3460426.3463636.
[5] D. Ma, F. Gao, Y. Bai, Y. Lou, S. Wang, T. Huang, L.-Y. Duan, From part to whole: who is behind the painting? in: Proceedings of the 25th ACM International Conference on Multimedia, 2017, pp. 1174–1182, doi:10.1145/3123266.3123325.
[6] E. Cetinic, T. Lipic, S. Grgic, Fine-tuning convolutional neural networks for fine art classification, Expert Syst. Appl. 114 (2018) 107–118, doi:10.1016/j.eswa.2018.07.026.
[7] M.K. Hosain, T.B. Taher, M.M. Rahman, et al., Genre recognition of artworks using convolutional neural network, in: 2020 23rd International Conference on Computer and Information Technology (ICCIT), IEEE, 2020, pp. 1–5, doi:10.1109/ICCIT51783.2020.9392688.

[8] E.L. Spratt, Dream formulations and deep neural networks: humanistic themes in the iconology of the machine-learned image, arXiv preprint arXiv:1802.01274(2018).
[9] A. Elgammal, B. Liu, D. Kim, M. Elhoseiny, M. Mazzone, The shape of art history in the eyes of the machine, in: 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, 2018, p. 21832191.
[10] L. Pilka, Digital curator, 2022, Accessed: 2 May 2024, https://www.digitalcurator.art.
[11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255, doi:10.1109/CVPR.2009.5206848.
[12] J. Horakova, J. Schimmel, P. Sikora, S. Miklanek, D. Barok, Vasulka live archive website, 2021, Accessed: 2 May 2024, https://vasulkalivearchive.net. 10.57734/v2mj-pe27
[13] P. Weibel, The apparatus world-a world unto itself: Eigenwelt der apparatewelt, in: Eigenwelt der Apparate-Welt. Pioneers of Electronic Art, Ars Electrocnica, 1992, p. 15-19.
[14] P. Weibel, An annotated Videography 1969-1976, Österreichischen Hochschúlerschaft, Innsbruck, 1977.
[15] P. Sikora, Right convolutional neural network for classification illustrations in artworks, in: Proceedings II of the Conference Student EEICT, 2021, p. 591-595.
[16] V. Sze, Y.-H. Chen, T.-J. Yang, J.S. Emer, Efficient processing of deep neural networks: a tutorial and survey, Proc. IEEE 105 (12) (2017) 2295–2329.
[17] S. Vasulka, Summer salt, 1982, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Summer%20Salt%20(video).
[18] S. Vasulka, Lava and moss, 2000, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Lava%20and%20Moss%20[Hraun%20og%20Mosi]%20(installation).
[19] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258, doi:10.1109/CVPR.2017.195.
[20] B. Zoph, V. Vasudevan, J. Shlens, Q.V. Le, Learning transferable architectures for scalable image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8697–8710, doi:10.1109/CVPR.2018.00907.
[21] C. Szegedy, S. Ioffe, V. Vanhoucke, A.A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-first AAAI Conference on Artificial Intelligence, 2017, pp. 4278–4284, doi:10.1609/aaai.v31i1.11231.
[22] A. Krizhevsky, G. Hinton, et al., Learning Multiple Layers of Features from Tiny Images, Citeseer, 2009.
[23] J.F. Gemmeke, D.P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R.C. Moore, M. Plakal, M. Ritter, Audio set: an ontology and human-labeled dataset for audio events, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2017, pp. 776–780.
[24] J. Pons, X. Serra, Randomly weighted CNNs for (music) audio classification, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2019, pp. 336–340, doi:10.1109/ICASSP.2019.8682912.
[25] J. Sang, S. Park, J. Lee, Convolutional recurrent neural networks for urban sound classification using raw waveforms, in: 2018 26th European Signal Processing Conference (EUSIPCO), IEEE, 2018, pp. 2444–2448, doi:10.23919/EUSIPCO.2018.8553247.
[26] Z. Zhang, S. Xu, S. Zhang, T. Qiao, S. Cao, Attention based convolutional recurrent neural network for environmental sound classification, Neurocomputing 453 (2021) 896–903, doi:10.1016/j.neucom.2020.08.069.
[27] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, M.D. Plumbley, PANNs: large-scale pretrained audio neural networks for audio pattern recognition, IEEE/ACM Trans. Audio Speech Lang. Process. 28 (2020) 2880–2894, doi:10.1109/TASLP.2020.3030497.
[28] S. Vasulka, W. Vasulka, 1-2-3-4, 1973, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/1-2-3-4%20(video).
[29] S. Vasulka, Bent scans, 2002, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Bent%20Scans%20(interactive%20installation).
[30] Z. Mushtaq, S.-F. Su, Efficient classification of environmental sounds through multiple features aggregation and data enhancement techniques for spectrogram images, Symmetry 12 (11) (2020) 1822, doi:10.3390/sym12111822.
[31] A.F. Agarap, Deep learning using rectified linear units (relu), arXiv preprint arXiv:1803.08375(2018).
[32] Meta AI, Logsoftmax — pytorch 1.12 documentation, 2022a, Accessed: 2 May 2024, https://pytorch.org/docs/stable/generated/torch.nn.LogSoftmax.html.
[33] Meta AI, Nllloss — pytorch 1.12 documentation, 2022b, Accessed: 2 May 2024, https://pytorch.org/docs/stable/generated/torch.nn.NLLLoss.html.
[34] S. Vasulka, Violin power, 1978, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Violin%20Power%20(video).
[35] S. Vasulka, W. Vasulka, Telc, 1974, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Telc%20(video).
[36] S. Pandit, S. Gupta, et al., A comparative study on distance measuring approaches for clustering, Int. J. Comput. Sci. Res. 2 (1) (2011) 29–31.
[37] S. Carter, Z. Armstrong, L. Schubert, I. Johnson, C. Olah, Activation Atlas, Distill, 2019, doi:10.23915/distill.00015. https://distill.pub/2019/activation-atlas
[38] W. Vasulka, Swan lake, 1971, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Swan%20Lake%20(video).
[39] R. Cabello, Three.js – javascript 3d library., 2022, Accessed: 2 May 2024, https://threejs.org/.
[40] S. Vasulka, W. Vasulka, Techne and eros, 1999, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Techne%20and%20Eros.
[41] S. Vasulka, Machine vision, 1978, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Machine%20Vision%20(installation).
[42] S. Vasulka, Allvision, 1976, Accessed: 2 May 2024, http://www.vasulka.org/Steina/Steina_AllVision/AllVision.html.
[43] R. Krauss, Video: The Aesthetics of Narcissism, 1, The MIT Press, 1976.
[44] S. Vasulka, Lilith, 1987, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Lilith%20(video).
[45] W. Vasulka, Artifacts, 1980, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Artifacts%20(video).
[46] W. Vasulka, Art of memory, 1987, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/Art%20of%20Memory%20(video).
[47] W. Vasulka, Brotherhood, 1992–1998, Accessed: 2 May 2024, https://vasulkalivearchive.net/Video/Play/The%20Brotherhood%20(installation).