Computer Vision Methods and Technologies as an Extension to the Ballet Industry

Annika Hoag

November 25, 2024

1 Problem Explanation

As technology is becoming more intertwined with everyday life, it is also becoming a tool used for support across various fields for many different purposes. Computer vision is one area of technology that is starting to be used more for body movement and pose recognition as new capture technologies and GPUs are becoming more available. In the domain of ballet dance, there are two problems that computer vision is fit to solve. The first is individual training correction, which comes from teachers needing to have their eyes on 10 or more students at a time during class. This can result in some dancers not receiving individual attention, which can then lead to incorrect technique, or even injuries, since in ballet the movements are made out of a strictly codified set of poses that must be performed correctly in order for the true ballet aesthetic to be shown. The other problem that arises in ballet dance is the accurate documentation of and assistance with choreography. Technology can essentially be used as a means to assist in ballet training [1].

2 Challenges

Despite the seemingly ideal scenario for new technological developments in the world of ballet, there are still some challenges with bringing these tools into this space.

[2] entailed the design and analysis of a research study that involved interviewing ballet teachers and pre-professional ballet dancers about their use and challenges with current technologies in the ballet industry. Their study found that there were a few reasons that most of the teachers and dancers considered themselves ”non-users” of current video technologies and tools. The first reason is that there is a limited benefit for teaching and learning ballet techniques with the assistance of these tools. Teachers and students limit their use of platforms like YouTube because they find it

Figure 1: Image of a typical ballet studio from [2]

difficult to find videos that represent good technique. Teachers prefer to use it for learning choreography from a specific video or to make their own YouTube videos for their students to practice choreography with. Some teachers in this study argued that using platforms like YouTube for technique training enforces the mimicry practice of just learning by copying, and that takes away the agency of the student and the ability for them to develop their own artistry. Ballet studios also usually do not have the right infrastructure to use technology tools. Traditionally, ballet studios only have tools like mirrors, special sprung floors, and barres (see Figure 1). Teachers have concerns about the way that using technology to learn technique will affect long-term learning outcomes. There is some concern, as well, because ballet pedagogy does not have a standardized approach when it comes to using technology.

In [2], it was said that dance may be the only field of education that has not been able to embrace technological applications because dancers and choreographers do not want a form of technological media to come between them and their live experience with dance. The researchers in [2] claim that it is important to augment the student to teacher relationship and experience instead of automating it. It is important to have a level of trust in the technologies, similar to the level of trust between a student and teacher. In general, dance technologies also have a low marketability and take longer to enter the marketplace. However, the teachers and dancers in this study believe that the real reason technology has not become commonplace in ballet education is because of the way that they are designed. Technology for the ballet industry should not be intrusive, but rather ”fade into the background" [2]. Financial barriers should also be considered when developing new ballet technologies since most dance studios struggle to stay in business as it is, and may not be able to afford these technologies.

Besides the challenges of accepting and implementing new technologies into the ballet world from a dancer perspective, there are also challenges for the developers of these technologies. There is a reasonable risk of noise in data collection, for example the dress of the dancer and the background they're performing in had to be carefully chosen in [3] so that skin color segmentation could work. A lot of the designs chosen by researchers involve the use of expensive hardware and other tools. This is creates a limitation where the work can only be done by the institutions that can afford them, creating another financial barrier. Ballet is also a form a of creative expression with some aesthetic interpretation unique to each dancer. This has the potential of making it difficult to always be able to determine the correctness of every pose or posture based solely on how similar it is compared to a supposed expert.

3 Review of Relevant Work

The researchers in [4] were looking for a tool to represent motion pattern using a hierarchical structure. They used segmental singular value decomposition (SegSVD) to represent motion patterns, both isolated motion classification and continuous motion. SegSVD is capable of capturing temporal information of time series and is good at matching patterns in a time series where the start and end points of the pattern are not previously known. The researchers created the *A-go-go* dance dataset for their work. They had an actor wearing a suit with 35 optical markers connected to various body parts which they used in creating the dataset. Their method of SegSVD outperformed the original SVD based method in the task of recognizing continuous dance sequences. They compared the average recognition accuracy results on four sets of test sequences against the results of continuous dynamic programming (CDP) and their SegSVD method outperformed CDP. Overall these researchers had higher recognition accuracy than previous works.

[3] proposed a fuzzy matching algorithm to automatically recognize unknown ballet postures from a set of 17 ballet dance positions. This was novel compared to other work in the field since most other work at the time dealt with more primitive postures like standing, bending, and crawling. The input to their algorithm was an unknown posture in RGB format. Skin color segmentation was then done on the image, followed by dilation. The output from these stages was used to make a minimized skeleton. The straight lines were sorted by their Euclidean distances, then were given fuzzy membership values with respect to each quadrant and compared with the 17 existing primitives. Figure 2 shows the progression of their algorithm with a known posture (ecarte devant) and an unknown image. This research resulted in an 82.35% recognition rate, recognizing 14 out of the

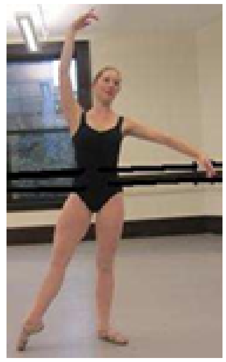
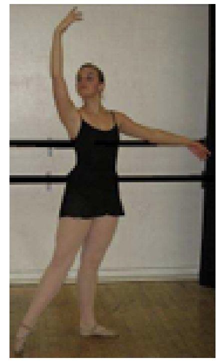
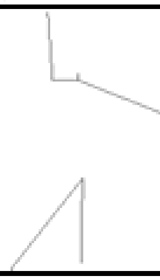
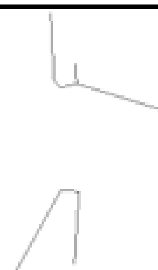
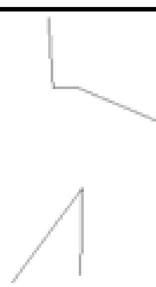
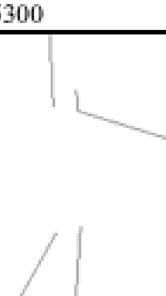
Features	Original Image	Unknown Image
RGB Image		
Skin Color Segmentation (after Dilation)		
Minimized Skeleton		
Approximated Straight Lines		
Length of all Straight Lines in descending order	98.0051, 74.0068, 73.0068, 60.1332, 21.0000, 6.0000	133.4541, 130.6484, 112.2185, 105.1190, 30.1496, 26.1725, 25.7099, 14.2127, 12.5300
Significant Straight Lines		

Figure 2: Comparison between original ecarté devant and an unknown image as seen in [3]

17 unknown postures correctly.

[5] developed a project (proof-of-concept) for tool that would help with choreography by recognizing and recording dance movements demonstrated by a choreographer. The researchers created and explored a program that could recognize a single ballet movement, and they used a Microsoft Kinect and Kinect SDK to track bodies and output joint locations. The researchers tested many different methods and algorithms for classification using the same set of training data (two dancers performing specific movements). These methods included L1 voting, L2 voting, linear support vector machine (SVM), quadratic SVM, and RBF (radial basis function) SVM. The accuracies for these methods were 80%, 71%, 55%, 56%, and 67% respectively, and the accuracies varied proportionally between the different dancers used in the data collection. The biggest issues the researchers ran into were in situations like the dancers using different leg than the other resulting in incorrect classification. The solution to this is simply having a more a complete library of reference. The researchers closed by suggesting an extension of their work in which the beats of music are identified and synchronizing dance movements to that beat to help choreographers with their musicality.

The researchers in [6] framed their problem from the perspective of beginner dancers, especially adult beginners. They claimed that dancers who have rehearsed the motions of choreographed movements know the sequences of them. They can recognize the movement feels (what the researchers referred to as ”embodied motion” [6]). After staging a ballet class for dancers who had never danced before, they found that fundamental positions are dependent on your own body and different for everyone. They determined that there is no one way to explain the ballet positions. They then decided to build a garment (see Figure 3) that would allow for natural changes in the relationship between students and teachers, and would allow for a more embodied form of feedback as the garment would be worn by the instructor.

The researchers wanted to create more of a connection from student to teacher. Their system would light up the limbs of the instructor and break each movement into starting and stopping positions. The team also had an assistant observing the class to watch for times where they weren't in unison, at which point they'd trigger lights on the instructor pointing out where the students should pay attention to. One limitation they found was durability because he circuits were joined with some hand tied knots. This was only the first stage of the garment, the researchers wanted to make some modifications and then conduct dance training experiments to make the final garment. [1] brought up another limitation of this technology which is the price and specialized construction necessary, claiming that a

Figure 3: A demonstration of the garment developed in [6]

vision-based approach would be more practical.

The problem to be solved in [7] was automatically identifying an unknown dance posture from 20 primitive postures of ballet. The researchers developed a new six stage algorithm to try and correctly identify these postures. Skin color segmentation was performed on the dance postures, and the output of that was dilated and processed to make stick figure skeletons of the original postures. Those stick figures were then modified to create minimized skeletons. The 20 postures were categorized into five groups based on their Euler number. This number as well as the line integral plots of the postures are what made up the initial database (group of unknown posture is found with Euler number and this posture's line integral plot is compared with the line integral plots of the postures in that group). There is a threshold that determines the correctness of the performed posture as well. Their proposed system had an accuracy of 91.35% in recognizing unknown postures. This work could be applied to assistance in ballet training by telling students what posture the teacher is using and/or providing feedback to the student regarding how correct their posture is compared to the teacher.

The work completed in [8] proposed a new framework for the real-time capture, assessment, and visualization of ballet movements performed by a student in an instructional virtual reality (VR) setting. This was made to provide a quantitative form of instructional feedback. The researchers set up a system with a CAVE (Cave Automatic Virtual Environment) that had four stereoscopic projectors and screens, and there was a tracking server that found each target's position and orientation based on captured images. That tracking data was what determined what content will be displayed on the screen. They used joint positional features and a proposed dance feature (found based on angles of joints relative to dancer's upper and lower torso) to create a spherical self-organizing map. This map was trained to find gestural trajectories that were then used as templates in the making of a library of predetermined dance movements to be used as an instructional set.

The group of authors for [8] measured recognition performance on a database of isolated gesture recordings done by teacher and student using L1 norm, L2 norm, and histogram intersection, providing average recognition rates from 90.5%-99.5%. There were different types of feedback given to the student visually in the CAVE in their proposed system side-by-side, overlay, and scoring. The students being instructed in the CAVE wore stereo glasses with optical markers to watch the performance of the virtual teacher to see her movements in 3D. They used experimental data to get the best teacher dance data and used that as a template for every gesture they were examining. The goal was to compare the student's dance performance to the teacher templates after the recognition stage. Students started to perform well (score of 95) by the sixth repetition of a movement, and the lowest scores were seen on the first time. At the time this paper was written the student performances were only showing moderate improvement since they were only using simple gestures

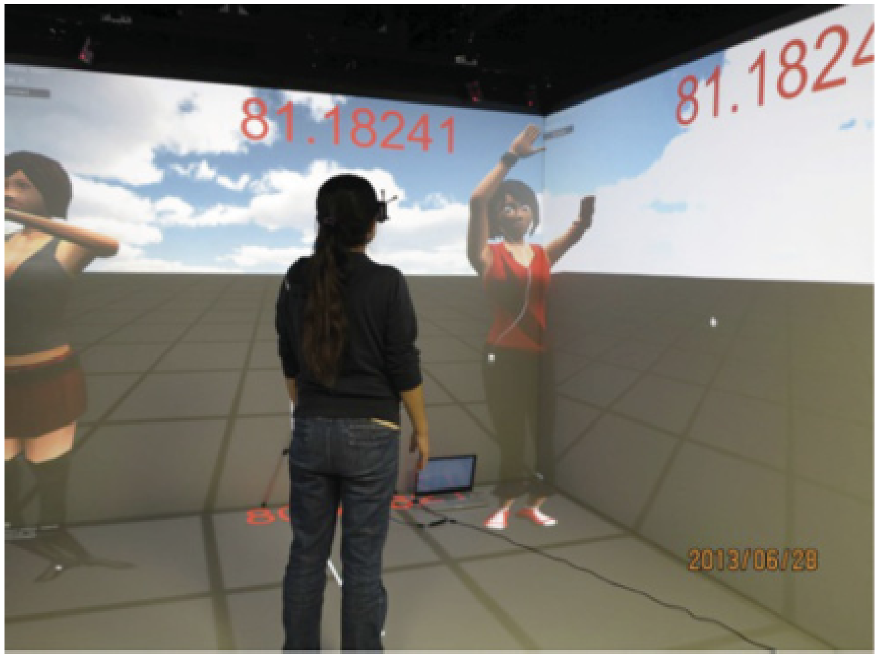


Figure 4: Example of student feedback with an overall score as seen in [8]

due to challenges in development, but there was slight improvement which they expected given the simple gestures.

The researchers in [1] attempted to solve the problem of individual training correction and choreographic assistance by examining and comparing different methods in computer vision. They implemented strict constraints in their data collection such as no mirrors or clutter in the image frames, good lighting, and having each dancer wear strict black ballet attire. The researchers used 30 dancers, showing eight ballet poses, and a Microsoft Kinect sensor and GoPro camera for the data collection. They analyzed three pipelines in their study to create the BaReCo model. The first pipeline was the traditional pipeline (Figure 4) which entailed capturing the dataset, pre-processing using grayscaling and histogram equalization localization (HOG method) to isolate the dancer's body in each frame, feature extraction (HOG also) to get key features to be in used in classification. The classification outputs were Support Vector Machine (SVM), Random Forest, and Gradient Boosted Tree (GBT). The second pipeline they used in [1] was the OpenPose pipeline. This was made up of the phases capturing, feature extraction, and classification. They captured their dataset as input into OpenPose's multi-stage convolutional neural network (CNN) architecture for feature extraction (find human skeleton key-point data from an image). The output of this pipeline were the classifications SVM, Rnadom Forect, or GBT, the saem as the traditional pipeline. The final pipeline the authors of [1] used was the VGG16 pipeline which used a VGG16 deep convolutional neural network (DCNN). They captured data as input to the DCNN, which was used

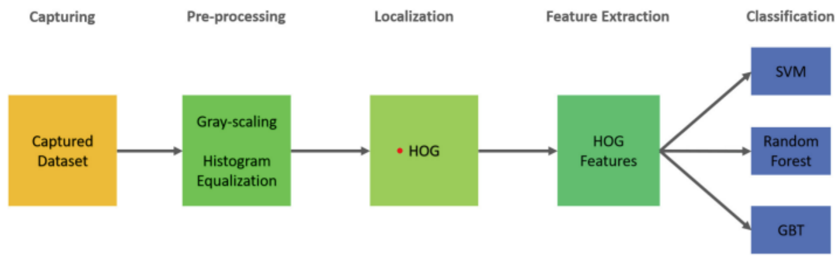


Figure 5: Traditional pipeline explored in [1]

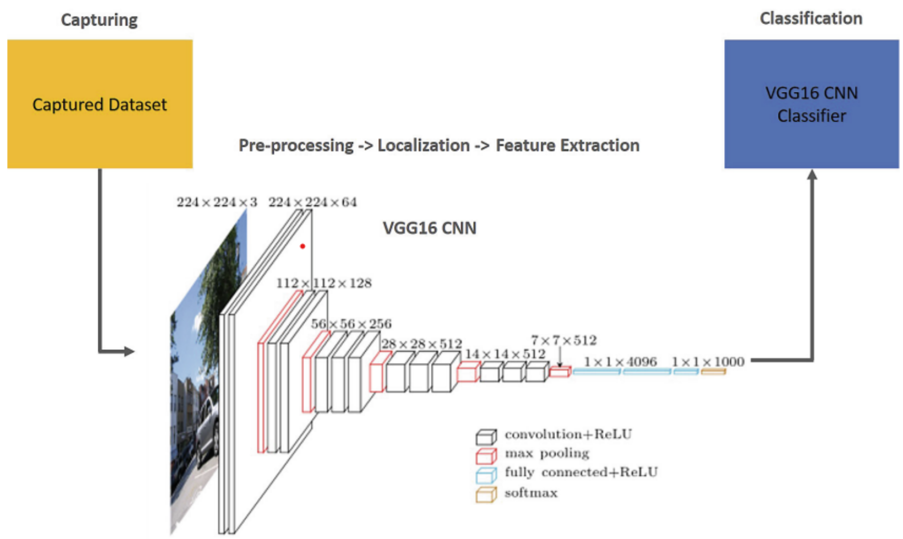


Figure 6: VGG16 pipeline architecture as seen in [1]

for pre-processing, localization, and feature extraction. The output was a VGG16 CNN Classifier. Figure 6 shows the architecture of the VGG16 pipeline. The researchers also utilized a family of algorithms called Region-Based CNNs (R-CNNs), one of these is called the Faster R-CNN which is more efficient than preceding algorithms since it uses Region Proposal Networks to find regions of interest in an image.

When comparing all of the pipelines the authors of [1] confirmed that OpenPose and Random Forest Approach achieved the best results when comparing the metrics of accuracy, equal error rate, recall, precision, specificity and F1 (recall & precision), but there was still accuracy scores above 95% for all classifiers. The researchers found that the poses that were incorrectly classified were mostly poses that were very similar to each other in the body. The major result of the paper was the validation of using computer vision in the ballet world to successfully complete the pose recognition task, specifically using the OpenPose approach and/or combining deep learning for feature extraction and traditional classifiers.

Another important contribution to the study of ballet's relationship with computer vision was [9]. This group worked on optimizing the design of motion recognition in the context of dance videos, with the goal of creating a computer that would have visual perception similar to that of a human. The videos in this study were preprocessed (grayscale conversion, background subtraction, and de-noising) to extract people from the images. Then the researchers used the Self Organized Map (SOM) approach to recognize the dance movements. The SOM approach is used in computer vision to reduce the execution cost of arithmetic. It is an unsupervised learning algorithm that does well in clustering and classifying data based on similarities between them because it preserves topological properties allowing for more similar movements to be grouped together with neighboring nodes. SOM is also good at handling noise and variability in data. They used SVM for motion recognition and compared the performance of hog3d and fast hog3d, testing the performance of SVM under many conditions. Results in simulations showed a 9.34% higher rate of accuracy than traditional arithmetic. This result comes from combining the gray channel with the inter-frame difference channel. The inter-frame difference channel was added to describe the areas with drastic changes in human movement. One limitation that was brought up is that their work only considered single dance movements so it didn't capture the full complexity of the dance art form. Their work did not factor in scene changes for instance.

The final relative work to be brought up here is in [10]. This paper used ballet to solve the more general problem of health training at home. The goal was to support the new use of ballet training to improve public health. The MediaPipe (from Google) was used to optimize video feedback. Trainers can use MediaPipe video feedback as an additional way to do "cloud movement" training in homes so that the trainers can reflect on issues and modify their training based on that. The researchers used the MediaPipe to estimate the coordinates of human joints, dividing the body into 33 points, in an image. The MediaPipe then uses machine learning to make pipelines to process the video data and provide feedback. With the plie movement, for instance, the stability of the experiments improved after three months of training with the MediaPipe video feedback.

4 Relevant Datasets

The authors of [11] provided a dataset [12] with their work. [11] gave a simple method for learning dance methods by imitation. The researchers took poses from the source subject and applied the learned pose-to-appearance mapping to make the target subjects. The authors provide a dataset made up of two parts, long videos of one dance that can be used both to train and evaluate their model. The dataset can be used for training motion transfer and methods of video generation.

A dataset called Let's Dance [13] was created by the authors of [14]. The dataset is made up of 1000 videos containing 10 visually similar dance categories that need motion for classification.

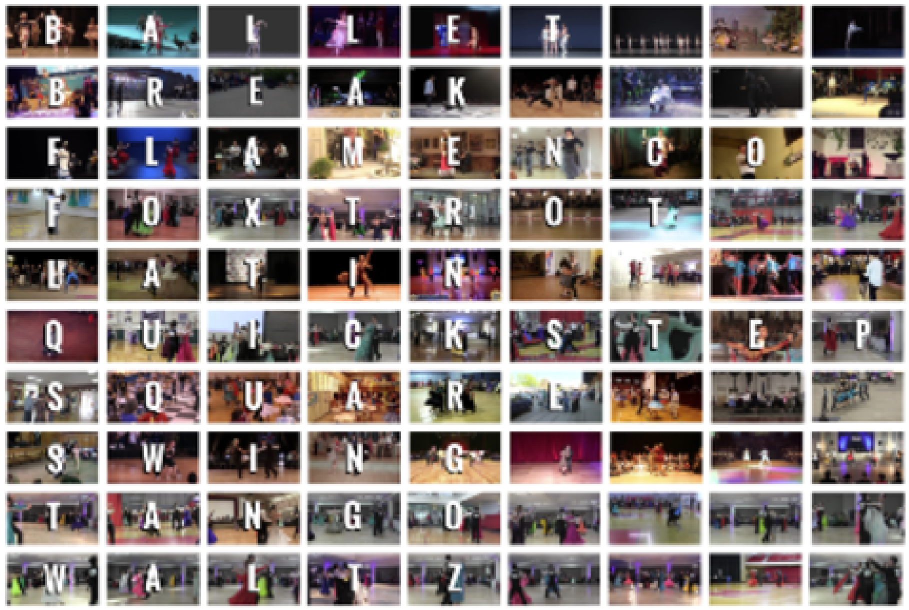


Figure 7: Visual representation of the dataset [13]

One of these categories is ballet which could be used for my final project. There are 100 videos here which could be useful as there are many frames to choose from out of 100 videos. Figure 7 is a visual representation of the dataset from the paper.

There is also a Kaggle dataset called ”Ballet Dancer” [15] with 24x24 images taken from ballet teaching videos. The other dance related datasets on Kaggle, unfortunately, are mostly related to hip hop, social dance, and Indian classical dance. Many of these dance datasets also fall into the realm of genre recognition, which is not as applicable to ballet pose recognition.

For this final project I will be using my own dataset that I created [16]. It contains images of three dance majors at Connecticut College doing eight standard ballet body positions, as well as frame by frame perspectives of them slowly transitioning between the poses. There are also videos of these transitions in the case that I need more images. The reason I am creating my own dataset is to simplify the data as there is little noise in the images and the pose is the main feature of the image so joint recognition should be simple in OpenPose. Creating my own dataset also allows for controlling exactly which poses I want to examine since the ballet world has a myriad of poses to use. Something like [15] could also be used as more testing data if there were enough images of the applicable poses. [13] could also be useful for an extension of this simple classification if we were to include video recognition or detection of these standard positions within choreography.

References

[1] Margaux Fourie and Dustin van der Haar. "Computer vision for the ballet industry: A comparative study of methods for pose recognition". In: *Business Information Systems: 23rd International Conference, BIS 2020, Colorado Springs, CO, USA, June 8–10, 2020, Proceedings 23*. Springer. 2020, pp. 118–129.

[2] Milka Trajkova and Francesco Cafaro. "Current use, non-use, and future use of ballet learning technologies". In: *Proceedings of the 2021 ACM Designing Interactive Systems Conference*. 2021, pp. 2052–2067.

[3] Sriparna Saha et al. "Fuzzy image matching for posture recognition in ballet dance". In: *2013 IEEE international conference on fuzzy systems (FUZZ-IEEE)*. IEEE. 2013, pp. 1–8.

[4] Liqun Deng et al. "Recognizing dance motions with segmental svd". In: *2010 20th International Conference on Pattern Recognition*. IEEE. 2010, pp. 1537–1540.

[5] Justin Dancs et al. "Recognition of ballet micro-movements for use in choreography". In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE. 2013, pp. 1162–1167.

[6] James Hallam et al. "Ballet hero: building a garment for memetic embodiment in dance learning". In: *Proceedings of the 2014 ACM International Symposium on Wearable Computers: Adjunct Program*. 2014, pp. 49–54.

[7] Sriparna Saha and Amit Konar. "Topomorphological approach to automatic posture recognition in ballet dance". In: *IET Image Processing* 9.11 (2015), pp. 1002–1011.

[8] Matthew Kyan et al. "An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 6.2 (2015), pp. 1–37.

[9] Yawen Pang and Yi Niu. "Dance video motion recognition based on computer vision and image processing". In: *Applied Artificial Intelligence* 37.1 (2023), p. 2226962.

[10] Miaowen Lei, Zuxuan Wang, and Fang Chen. "Ballet Form Training Based on MediaPipe Body Posture Monitoring". In: *Journal of Physics: Conference Series*. Vol. 2637. 1. IOP Publishing. 2023, p. 012019.

[11] Caroline Chan et al. "Everybody dance now". In: *Proceedings of the IEEE/CVF international conference on computer vision*. 2019, pp. 5933–5942.

[12] Caroline Chan et al. *Everybody Dance Now*. `https://carolineec.github.io/everybody_dance_now/`. [Online; accessed November 24, 2024].

[13] Daniel Castro et al. *Let's Dance: Learning From Online Dance Videos*. https://sites.cc.gatech.edu/cpl/projects/dance/. [Online; accessed November 24, 2024].

[14] Daniel Castro et al. "Let's dance: Learning from online dance videos". In: *arXiv preprint arXiv:1801.07388* (2018).

[15] *Ballet Dancer*. https://www.kaggle.com/datasets/msliwang/ballet-dancer. [Online; accessed November 24, 2024].

[16] *images of Conn Dancers*. https://drive.google.com/drive/folders/12GpsXAIGU0nfozaoHe2fx6i [Online; accessed November 24, 2024].