A Dissertation Review Report

on

UNCERTAIN TASK PREDICTION USING CONTINUAL LEARNING

Submitted in partial fulfillment of the
requirement for the award of the degree of

BACHELOR OF TECHNOLOGY(CSE)

GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of
Dr. John A
Associate Professor

Submitted By

Suraj Kumar Singh (18SCSE1010647)

Abhishek Dubey (18SCSE1010427)

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA
INDIA

DECEMBER, 2021

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"**SEQUENCE OF ACTION RECOGITION USING CONTINUAL LEARNING**"** in partial fulfillment of the requirements for the award of the B.Tech submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of September, 2021 to December and 2021, under the supervision of Dr. John A, Assistant Professor, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Abhishek Dubey,Suraj Kr. Singh

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Dr. John A
Assistant Professor

CERTIFICATE

The Final Thesis/Project/ Dissertation Viva-Voce examination of Suraj Kumar Singh,Abhishek Dubey has been held on __________________ and his/her work is recommended for the award of Integrated B.Tech (CSE).

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: November, 2013 Place: Greater Noida

Abstract

Continual Learning is one of the founding stones of Artificial Intelligence and Machine Learning that will help it grow big in future. At present most of the AI and ML powered robots or software's do not follow such approach because of the complexity in application of such techniques in real life.

Continual Learning also known as 'Learning of Learning' or 'Lifelong Learning' is nothing but a way in which a machine will be able to reuse the knowledge it has to further decode more knowledge, just like humans do. It will be primarily achieved by Unsupervised Learning method of learning. Such models will continuously keep on improving over time without being explicitly programmed and won't require much of human interference. Using this technique we can even extract meaningful and reliable knowledge from an unreliable source. Continual Learning is the actual and the fundamental concept on which AI & ML are based upon. With this we can even achieve machine level intelligence. The benchmark to achieve here is to implement the sequential learning process. The major challenge in lifelong learning is to continue learning without catastrophic forgetting, which means a machine immediately or abruptly overlaps the previously learned knowledge and replaces it with the new one. Moreover another challenge in CL is that deployment of models should be taken care of, since it's not same as regular deployment so user experience and accuracy must be taken care of.

CL is very important because data trends keep on changing, new technologies and trends are changing the internet every now and then. So a system has to be there that will continuously keep on accepting the data and extract patterns and knowledge out of it.

Contents

List of Figures

Acronyms

AGI	Artificial General Intelligence
CL	Continual Learning
AI	Artificial Intelligence
ML	Machine Learning
DL	Deep Learning

CHAPTER-1

INTRODUCTION

Recognition of actions and its prediction is a crucial and a challenging task in video processing and action classification field. It has become basis for diverse applications, such as healthcare, elderly surveillance, child care, human position estimation, home monitoring etc. The use-cases are endless. Despite the important progress in human activity recognition gathered from video sequences, it remains a delicate problem for many reasons, such as changes in viewpoint and distance from the camera, the complexity of the background, and the diversity of speed.

The extraction of significant features is the most challenging part. Indeed, it influences the performance of the algorithm by reducing the time and complexity of calculations. However, the most traditional methods of human activity recognition [1] are based on handcrafted local features from RGB video taken by 2D cameras that are unable to manage complex activities. Some methods are based on the detection of a moving person by background extraction. Gaussian Background [2], kernel density estimation [3], Visual Background Extractor [4], and Sigma Delta [5] have been used with success in static fonts. Other methods involve motion tracking to characterize human activity. Human tracking is done to control its movement and build trajectories throughout the sequence. Tracking is usually a simple procedure for humans. The problem becomes complex when the speed of the objects is very high. This encourages researchers to develop methods that solve motion tracking problems using computer vision methods such as Optical Flow [6,7], Scale Invariant Feature Transformation (SIFT) [8], Histogram of Oriented Gradient (HOG) [9], and Mean Shift [10]. Since these approaches recognize actions based on the appearance and motion of parts of the human body from RGB video sequence, they lack a 3D structure from the scene.

Therefore, recognizing a human action based only on RGB modality is not enough to overcome current challenges. With the development of artificial intelligence and high computing capacity, some deep learning and transfer learning methods are adopted for the learning and automatic extraction of complex features provided by sensors.

Recently, the recognition of human actions based on the skeleton has attracted the attention of several researchers. Some action recognition approaches using RGB-D cameras and skeleton presentation have been proposed and have an advanced state-of-the-art status. Therefore, the recognition of actions and the analysis of human behaviors in an intelligent home or in an indoor position are becoming more and more important. Indeed, in the field of medicine, there is a rapidly increasing demand for systems to recognize human actions and to quickly detect patients' physical and mental health problems. The aim of this work was to propose a system for human activity prediction using continual learning based on the presentation of human features in a video sequence.

To address the challenges mentioned above, we propose a continual learning scheme that can improve human motion prediction accuracy while reducing the risk of catastrophic forgetting.

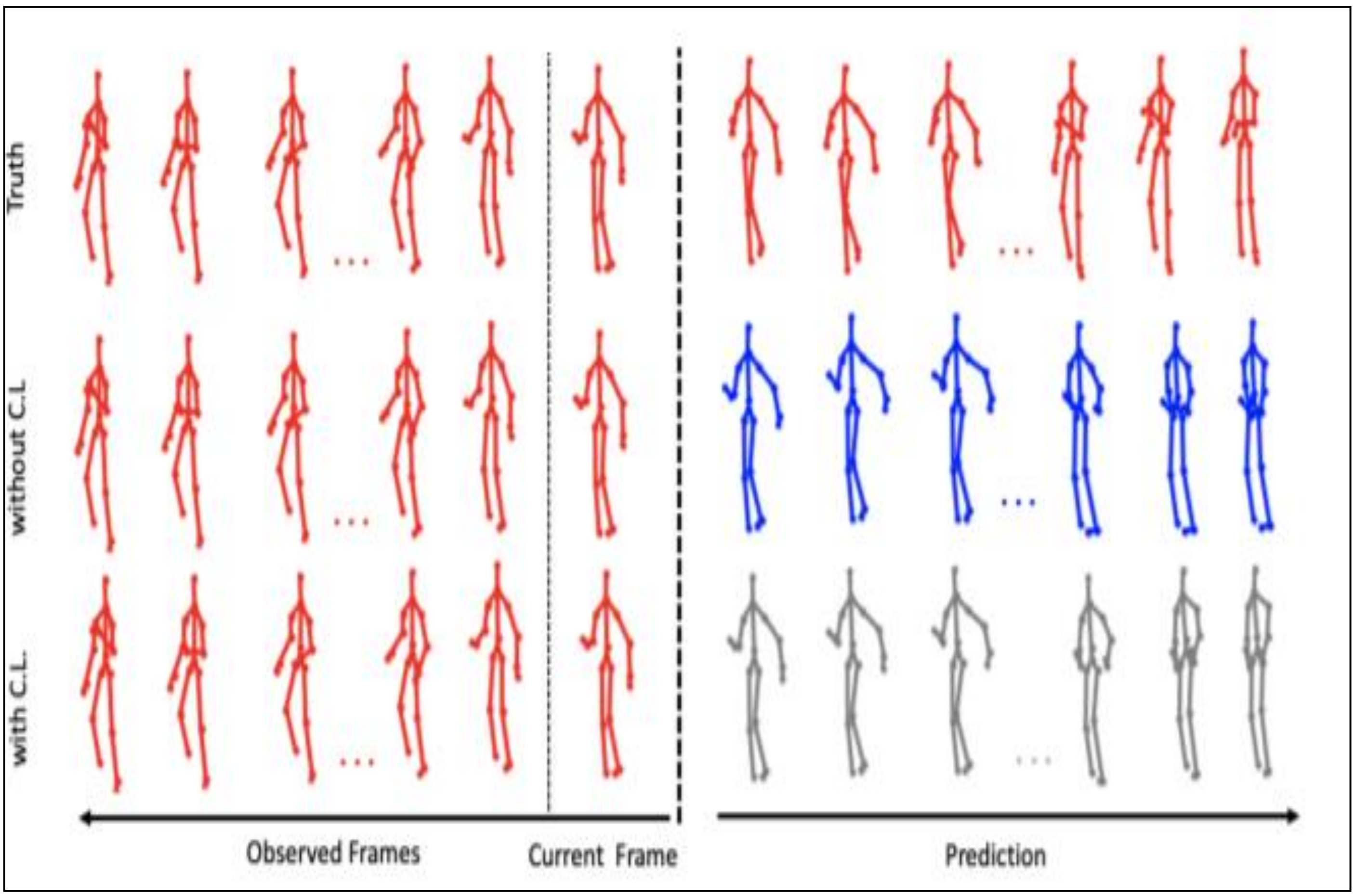


Fig 1. Qualitative performance of different motion prediction

1.1 Artificial Intelligence

Artificial intelligence is a wide-ranging branch of computer science concerned with building smart machines capable of performing tasks that typically require human intelligence. A.I includes a variety of branches in it, i.e., Machine Learning and Deep Learning are subsets of A.I.

According to Jhon McCarthy's definition coined in 1956, A.I is defined as "the science of making intelligent machines".

Artificial Intelligence is is mainly divided into two broad categories. It represented in Figure

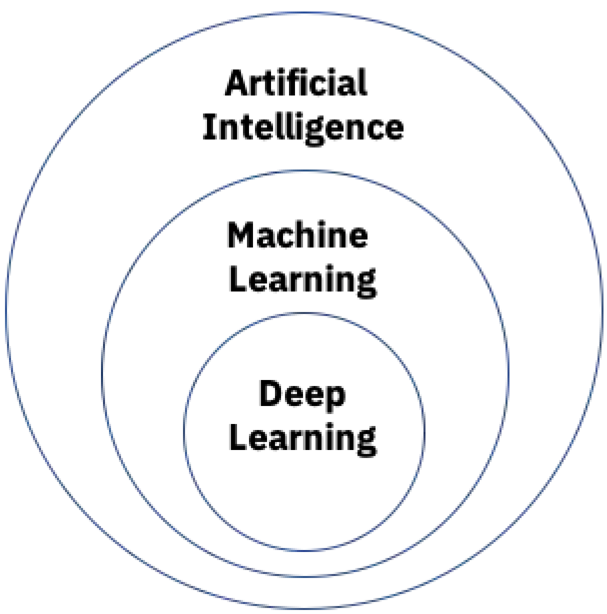


Fig 1 : Categories of AI

1.Narrow AI: Sometimes referred to as "Weak AI," this kind of artificial intelligence operates within a limited context and is a simulation of human intelligence. Narrow AI is often focused on performing a single task extremely well and while these machines may seem intelligent, they are operating under far more constraints and limitations than even the most basic human intelligence.

2.Artificial General Intelligence: AGI, sometimes referred to as "Strong AI," is the kind of artificial intelligence we see in the movies, like the robots from *Westworld* or Data from *Star Trek: The Next Generation*. AGI is a machine with general intelligence and, much like a human being, it can apply that intelligence to solve any problem. Narrow AI is all around us and is easily the most successful realization of artificial intelligence to date.

1.2 Machine Learning

Machine learning is a subset of artificial intelligence that gives systems the ability to learn and optimize processes without having to be consistently programmed. Simply put, machine learning uses data, statistics and trial and error to “learn” a specific task without ever having to be specifically coded for the task.

Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, uncovering key insights within data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase, requiring them to assist in the identification of the most relevant business questions and subsequently the data to answer them.

There are mainly three types of machine learning techniques:

1. **Supervised ML:** Supervised learning, also known as supervised machine learning, is defined by its use of labeled datasets to train algorithms that to classify data or predict outcomes accurately.
2. **Semi-Supervised ML**: Semi-supervised learning offers a happy medium between supervised and unsupervised learning. During training, it uses a smaller labeled data set to guide classification and feature extraction from a larger, unlabeled data set.
3. **Unsupervised ML:** Unsupervised learning, also known as unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabeled datasets.

Some real world use cases of Machine Learning include:

1. **Speech recognition.**
2. **Customer service.**
3. **Computer vision.**
4. **Recommendation engines.**
5. **Automated stock trading**

1.3 Deep Learning

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. These neural networks attempt to simulate the behavior of the human brain albeit far from matching its ability allowing it to learn from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy.

Deep learning neural networks, or artificial neural networks, attempts to mimic the human brain through a combination of data inputs, weights, and bias. These elements work together to accurately recognize, classify, and describe objects within the data. Deep neural networks consist of multiple layers of interconnected nodes, each building upon the previous layer to refine and optimize the prediction or categorization. This progression of computations through the network is called forward propagation. The input and output layers of a deep neural network are called visible layers. The input layer is where the deep learning model ingests the data for processing, and the output layer is where the final prediction or classification is made. Another process called backpropagation uses algorithms, like gradient descent, to calculate errors in predictions and then adjusts the weights and biases of the function by moving backwards through the layers in an effort to train the model. Together, forward propagation and backpropagation allow a neural network to make predictions and correct for any errors accordingly. Over time, the algorithm becomes gradually more accurate.

Some real world use cases of Deep Learning include:

1. Law Enforcement
2. Financial Services
3. Chatbots
4. HealthCare

SOFTWARE REQUIREMENT :

As the project is developed in python, we have used Anaconda for Python 3.6.5 and Spyder.

1. Anaconda

It is a free and open source distribution of the Python and R programming languages for data science and machine learning related applications (largescale data processing, predictive analytics, scientific computing), that aims to simplify package management and deployment. Package versions are managed by the package management system conda. The Anaconda distribution is used by over 6 million users, and it includes more than 250 popular data science packages suitable for Windows, Linux, and MacOS.

2. Spyder

Spyder (formerly Pydee) is an open source cross-platform integrated development environment (IDE) for scientific programming in the Python language. Spyder integrates NumPy, SciPy, Matplotlib and IPython, as well as other open source software. It is released under the MIT license. Spyder is extensible with plugins, includes support for interactive tools for data inspection and embeds Python-specific code quality assurance and introspection instruments, such as Pyflakes, Pylint and Rope. It is available cross-platform through Anaconda, on Windows with WinPython and Python 6 (x,y), on macOS through MacPorts, and on major Linux distributions such as Arch Linux, Debian, Fedora, Gentoo Linux, openSUSE and Ubuntu.

❖ **Features include:**

- editor with syntax highlighting and introspection for code completion
- support for multiple Python consoles (including IPython)
- the ability to explore and edit variables from a GUI

❖ **Available plugins include:**

- Static Code Analysis with Pylint
- Code Profiling
- Conda Package Manager with Conda

❖ **Hardware Interfaces**

1. Processor : Intel CORE i5 processor with minimum 2.9 GHz speed.
2. RAM : Minimum 4 GB.
3. Hard Disk : Minimum 500 GB

Software Interfaces

1. Microsoft Word 2003 7 2. Database
2. Storage : Microsoft Excel
3. Operating System : Windows10 1.3

Motivation:-

In previous time, for psychologist, analyzing facial expression was an essential part. Nowadays image processing have motivated significantly on research work of automatic face mood detection. There are lots of depressed people lived in our society. Also lots of busy people those who do not know their present mental condition. So we try to develop such an application and by this application they will able to see their present mental condition

- ❖ **Continual Learning**

Continual Learning is built on the idea of *learning continuously and adaptively about the external world* and enabling the *autonomous incremental development* of ever more complex *skills* and *knowledge*.

In the context of *Machine Learning* it means being able to *smoothly update the prediction model* to take into account different tasks and data distributions but still being able to **re-use** and **retain** useful knowledge and skills during time. Hence, **CL** *is the only paradigm* which force us to deal with an higher and realistic time-scale where data becomes available only during time, we have no access to previous perception data and it's imperative to build on top of previously learned knowledge. Examples of countinal learning shown in the figure

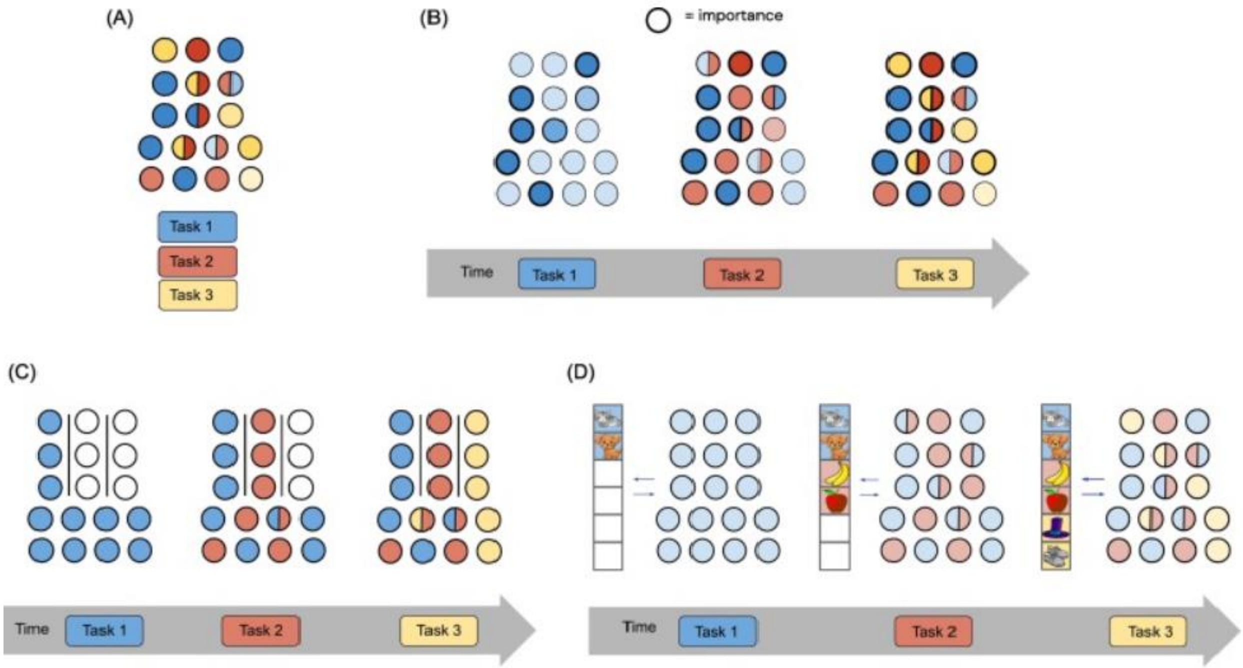


Fig : Example of countinal learning

❖ Applications of Continual Learning

Continual Learning is a concept to learn a model for a large number of tasks sequentially without forgetting knowledge obtained from the preceding tasks, where the data in the old tasks are not available any more during training new ones.

1. Senitment Analysis: It is one of the beneficial and real-time machine learning applications that help determine the emotion or opinion of the speaker or the writer.
2. Virtual Assistants: Continual Learning can be implemented in Virtual Assistants or bots so as to improve performance and human intervention required dramatically.
3. Product Recommendation in E-Commerce: Products keep on changing and new products keep on coming over time with continual learning explicit programming and training new models.
4. Sequence Analysis: With continual learning and unsupervised learning multiple sequence analysis techniques can be implemented, which can be helpful in creating many Machine Learning models which are exceptionally good.

By contrast, our brains work in a very different way. We are able to learn incrementally, acquiring skills one at a time and applying our previous knowledge when learning new tasks.

Computational models of intrinsic motivation have taken inspiration from the way human infants and children choose their goals and progressively acquire skills to define developmental structures in lifelong learning frameworks[24].

Multi-task learning, transfer learning, and related methods have a long history.

In brief, our Learning without Forgetting approach could be seen as a combination of Distillation Networks [26] and fine-tuning [25]. Fine-tuning initializes with parameters from an existing network trained on a related data-rich problem and finds a new local minimum by optimizing parameters for a new task with a low learning rate.

Less Forgetting Learning [27] is also a similar method, which preserves the old task performance by discouraging the shared representation to change. This method argues that the task-specific decision boundaries should not change, and keeps the old task's final layer unchanged, while our method discourages the old task output to change, and jointly optimizes both the shared representation and the final layer. We empirically show that our method outperforms Less Forgetting Learning on the new task [10].

LEARNINGWITHOUTFORGETTING:

Start with:

θ_s: shared parameters

θ_o: task specific parameters for each old task

X_n, Y_n: training data and ground truth on the new task

Initialize:

$Y_o \leftarrow \text{CNN}(X_n, \theta_s, \theta_o)$ *// compute output of old tasks for new data*

$\theta_n \leftarrow \text{RANDINIT}(|\theta_n|)$ *// randomly initialize new parameters*

Train:

Define $\hat{Y}_o \equiv \text{CNN}(X_n, \hat{\theta}_s, \hat{\theta}_o)$ *// old task output*

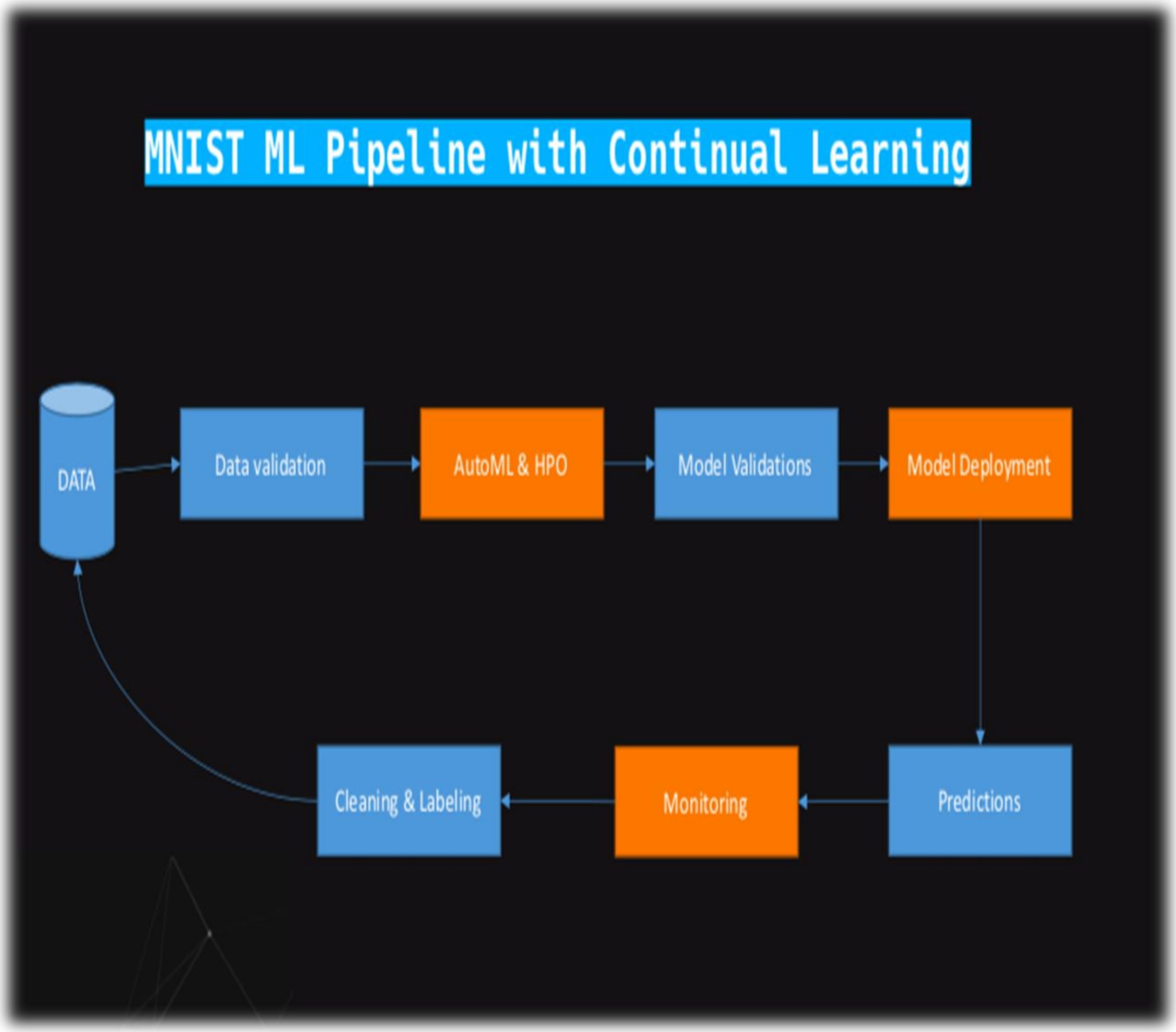
Define $\hat{Y}_n \equiv \text{CNN}(X_n, \hat{\theta}_s, \hat{\theta}_n)$ *// new task output*

$$\theta_s^*,\ \theta_o^*,\ \theta_n^* \leftarrow \underset{\hat{\theta}_s, \hat{\theta}_o, \hat{\theta}_n}{\operatorname{argmin}} \left(\lambda_o \mathcal{L}_{old}(Y_o, \hat{Y}_o) + \mathcal{L}_{new}(Y_n, \hat{Y}_n) + \mathcal{R}(\hat{\theta}_s, \hat{\theta}_o, \hat{\theta}_n) \right)$$

Fig: Learning Without Forgetting [24].

Block Diagram for Continual Learning

FIG: CONTINUAL LEARNING BLOCK DIAGRAM



❖ PROBLEM DEFINITION

Using Continual Learning we are trying to approach and solve a real world problem of action recognition. Currently, AI ML and DL are facing several challenges and difficulties, mainly attributed to the extraction of information in real-time from a large number. The extracted information can be useful to identify and detect many events that can help in many analyses, such as abnormal events and people's behavior, as well as to predict events that usually happen in the scenes.With the use of continual learning we will be able to overcome such problems of catastrophic forgetting, and produce a better model for real world usecase.

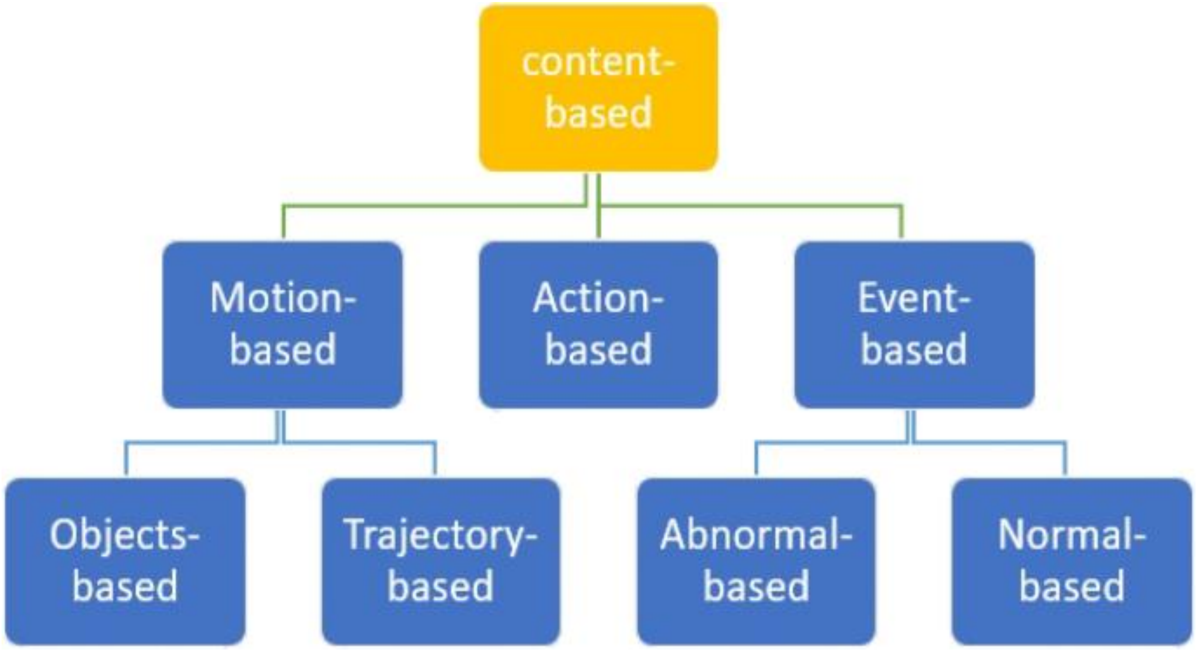


Fig: Flowchart for classification of solution approaches

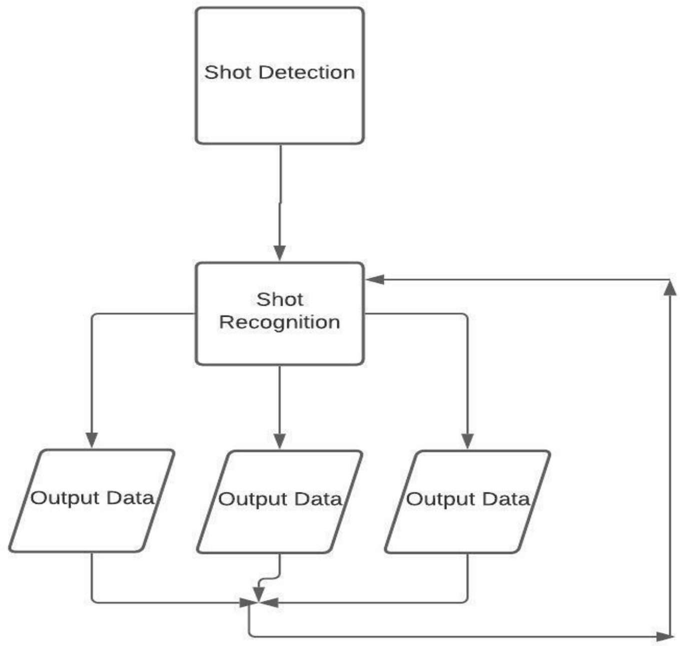
❖ PROPOSED WORK

We propose a new approach for a combined multiple human action recognition and summarization technique. In our method, we start with detecting human bodies using a proposed background-subtraction-based approach. Then, each one was tracked separately to generate a corresponding video sequence of each person during his presence in the scene. We designed the training part to represent all categories of human actions by a set of Histograms of Oriented Gradient of the Temporal Difference Map that represents the motion history of the target.

After the recognition stage, we will try to summarise the actions from the captured timeline and predict the output.

In current implementation scenario we have,y = f(x) where y is the dataset and f(x) is the data obtained. The problem here is y is always fixed and does not change or improve with time. However with continual learning or life long learning method, the y will also change and reduce human intervention to a great extent.

Moreover with implementation of continual learning into AI ML and DL, all these technologies will reach towards a real meaning of what intelligence actually is. The proposed work black diaram shown in the figure



CHAPTER-2

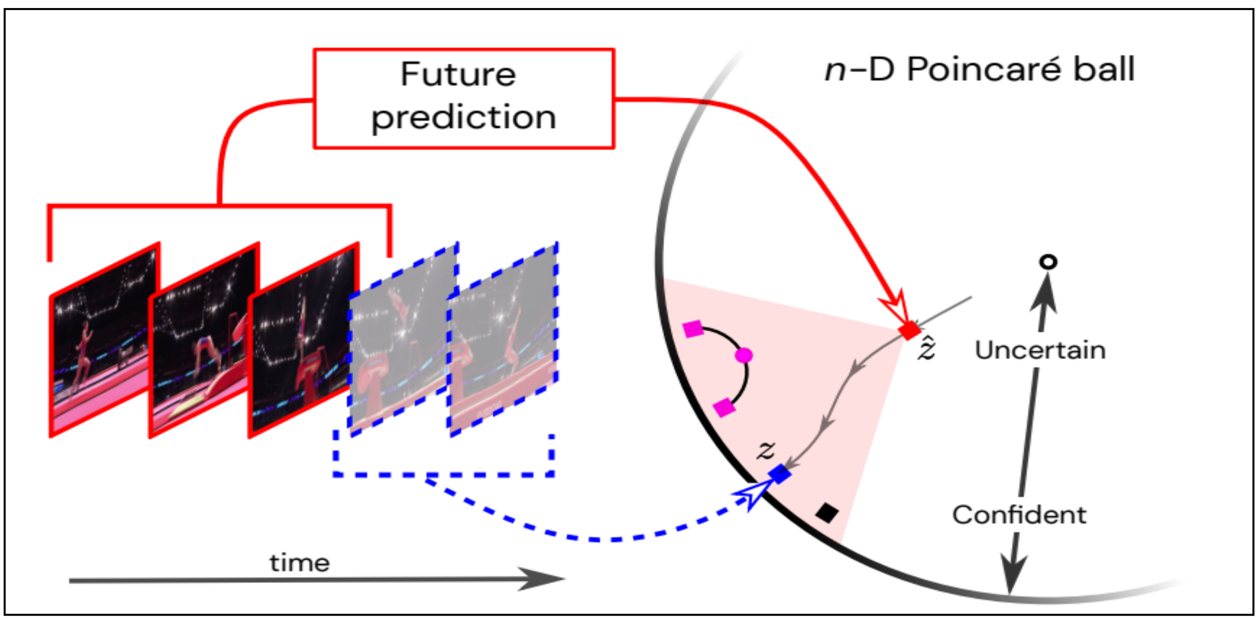
RELATED WORK

Video representation learning aims to learn strong features for a number of visual video tasks. By taking advantage of the temporal structure in video, a line of work uses the future as incidental supervision for learning video dynamics. Since generating pixels is challenging, instead use the natural temporal order of video frames to learn self-supervised video representations. Similar to self-supervised image representations, temporal context also provides strong incidental supervision . A series of studies from Oxford has investigated how to learn a representation in the future using a contrastive objective. We urge readers to read these papers in detail as they are the most related to our work. While these models learn predictable features, the underlying representation is not adaptive to the varying levels of uncertainty in natural videos. They also focus on action recognition, and not action prediction. By representing action hierarchies in hyperbolic space, our model has robust inductive structure for hedging uncertainty.

Unlike action recognition, future action prediction and early action prediction are tasks with an intrinsic uncertainty caused by the unpredictability of the future. Future action prediction infers future actions conditioned on the current observations. Approaches for future prediction range from classification, through prediction of future features, to generation of the future at the skeletal or even pixel levels. Early action prediction aims to recognize actions before they are completely executed. While standard action recognition methods can be used for this task, most approaches mimic a sequential data arrival. We evaluate our self-supervised learned representations on these two tasks.

Hyperbolic embedding has emerged as excellent hierarchical language representations in natural language processing. These works are pioneering. Riemannian optimization algorithms are used to optimize the models using hyperbolic geometry. Their success is largely attributed to the advantage of hyperbolic space to represent hierarchical structure. Following the Poincare embedding use a hyperbolic entailment cone to represent the hierarchical relation in an acyclic graph. Further applies hyperbolic geometry to feed forward neural networks and recurrent neural networks. Since visual data is naturally hierarchical, hyperbolic space provides a strong inductive bias for images and videos as well. Perform several image tasks, demonstrating the advantage of hyperbolic embeddings over Euclidean ones. proposes video and action embeddings in the hyperbolic space and trains a cross-modal model to perform hierarchical action search. We instead use hyperbolic embeddings for prediction and we use the hierarchy to model uncertainty in the future. We also learn the hierarchy from self-supervision, and our experiments show an action hierarchy emerges automatically. Since dynamics are often stochastic, uncertainty representation underpins predictive visual models. There is extensive work on probabilistic models for visual prediction, and we only have space to briefly review. For example, measures the covariance between outputs generated under different dropout masks, which reflects how close the predicted state is to the training data manifold. A high covariance indicates that the model is confident about its prediction. use assembling to estimate uncertainty, grounded in the observation that a mixture of neural networks will produce dissimilar predictions if the data point is rare. Another line of work focuses on generating multiple possible future prediction, such as variation auto-encoders (VAE) variation recurrent neural networks (VRNN) and adversarial variations. These models allow sampling from the latent space to capture the multiple outcomes in the output space. Probabilistic approaches are compatible with our framework. The main novelty of our method is that we represent the future uncertainty hierarchically in a hyperbolic space. The hierarchy naturally emerges during the process of learning to predict the future.

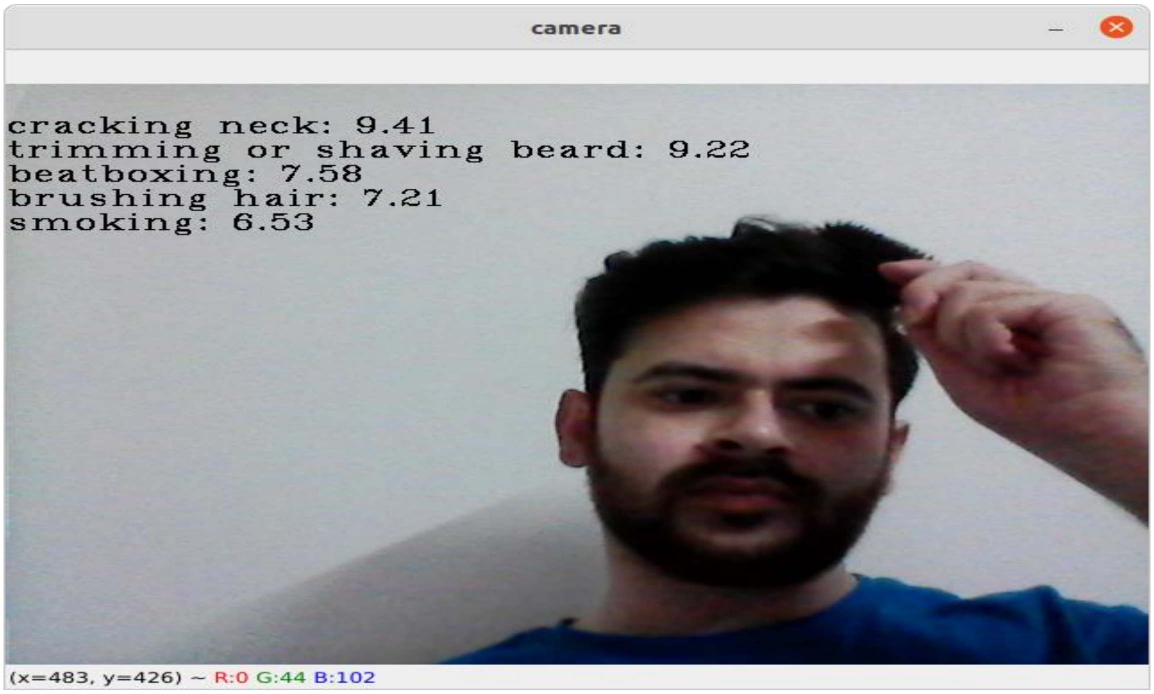
Fig 2. The future is non-deterministic. Given a specific past, different representations can encode different futures, all of them possible. In case the model is uncertain, it will predict an abstraction of all these possible futures, represented by zˆ. The more confident it is, the more specific the prediction can get.



CHAPTER 3

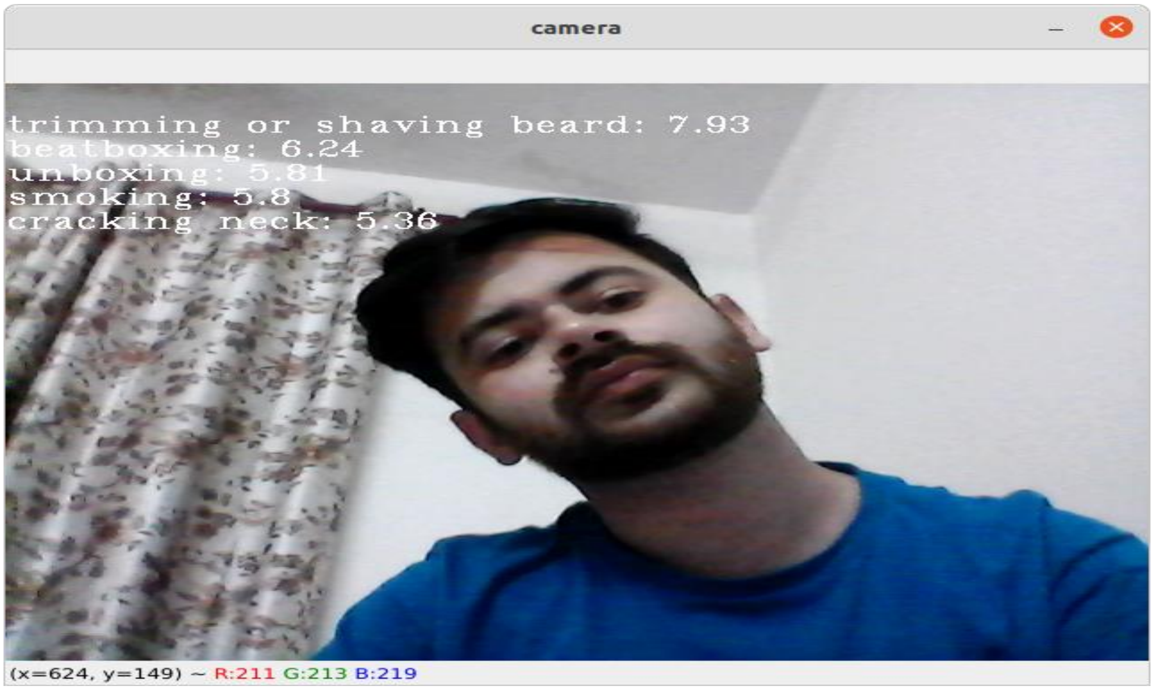
Screenshots of Sample Outputs

1.



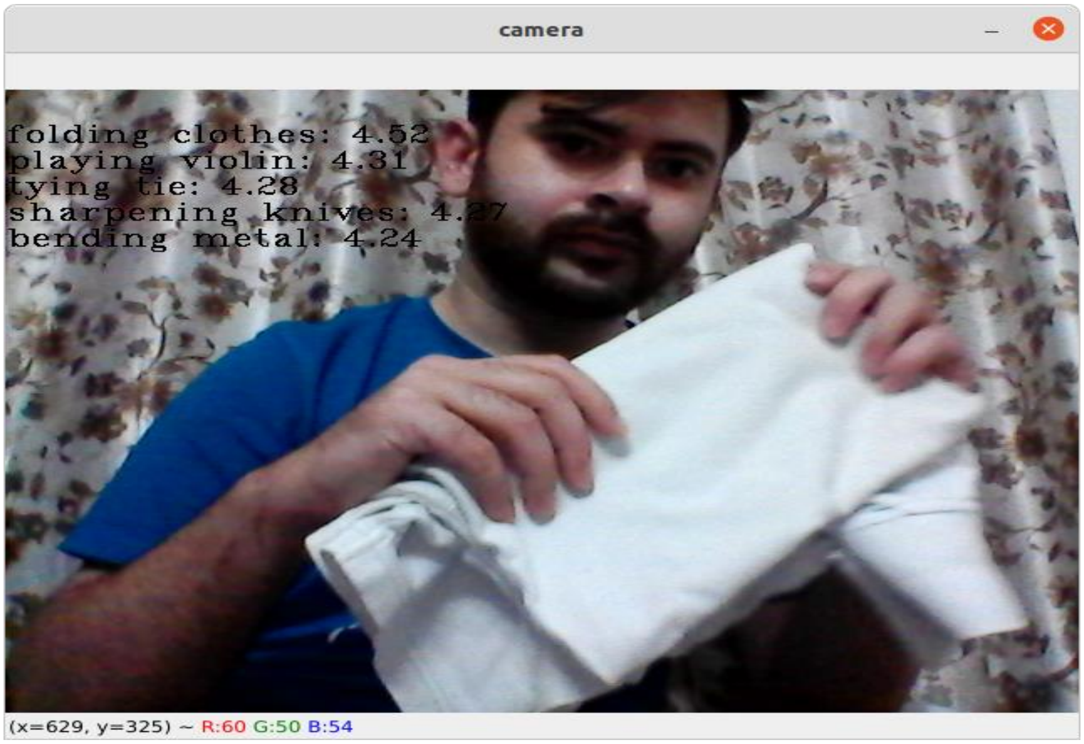
BRUSHING HAIR

2.



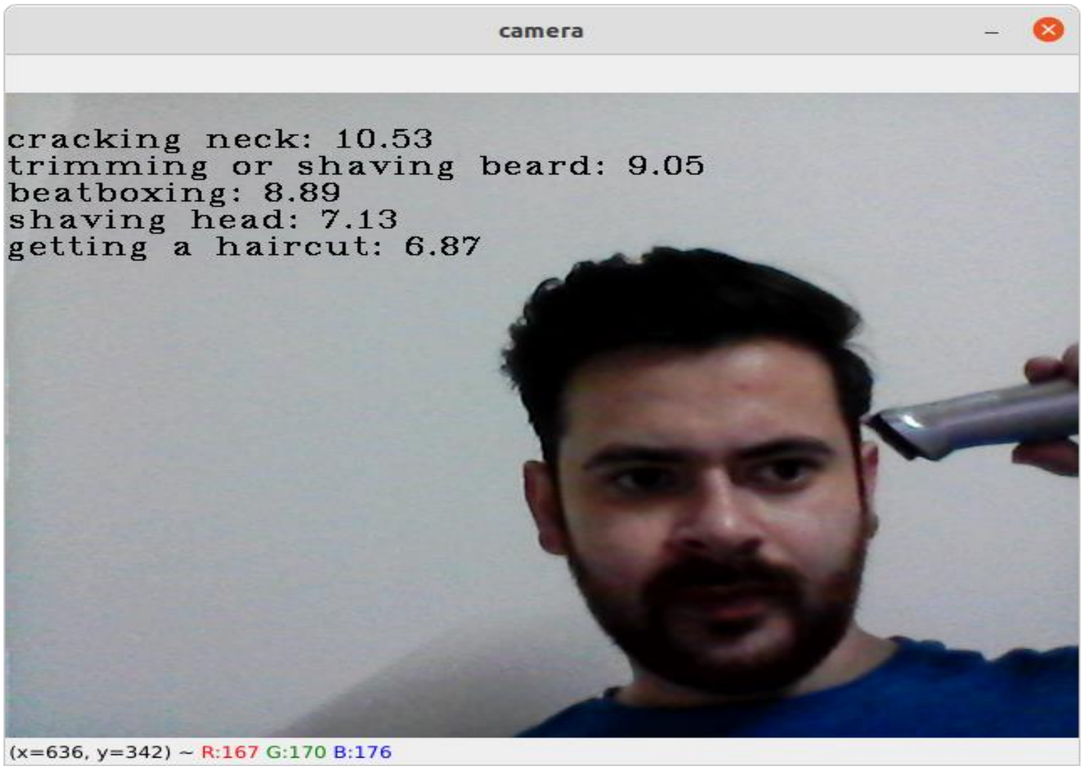
CRACKING NECK

3.



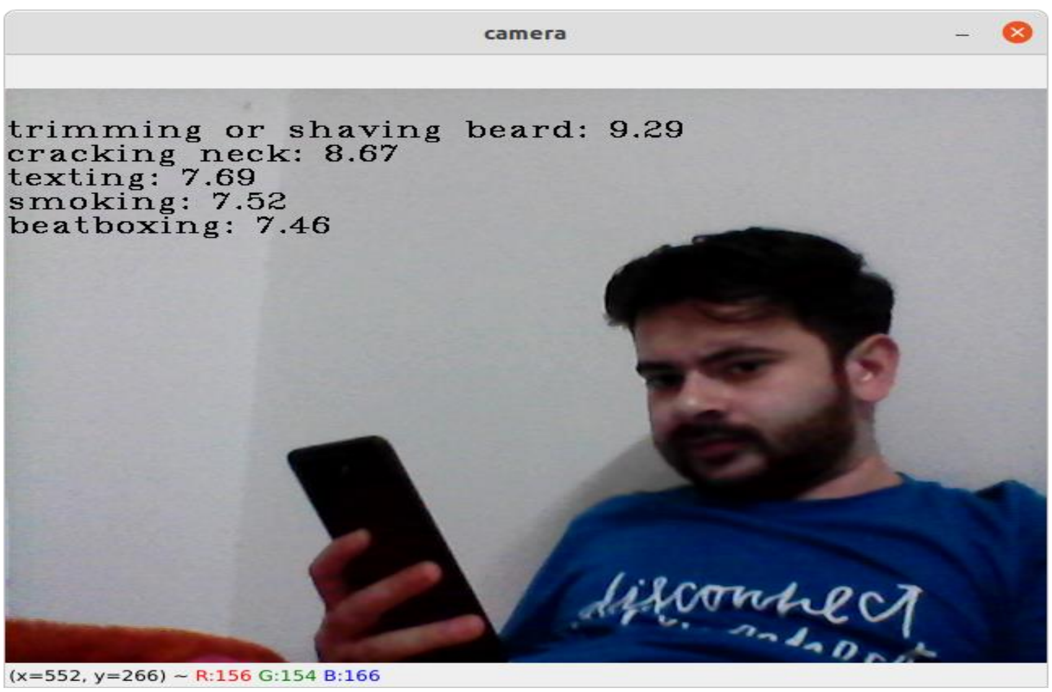
FOLDING CLOTHES

4.



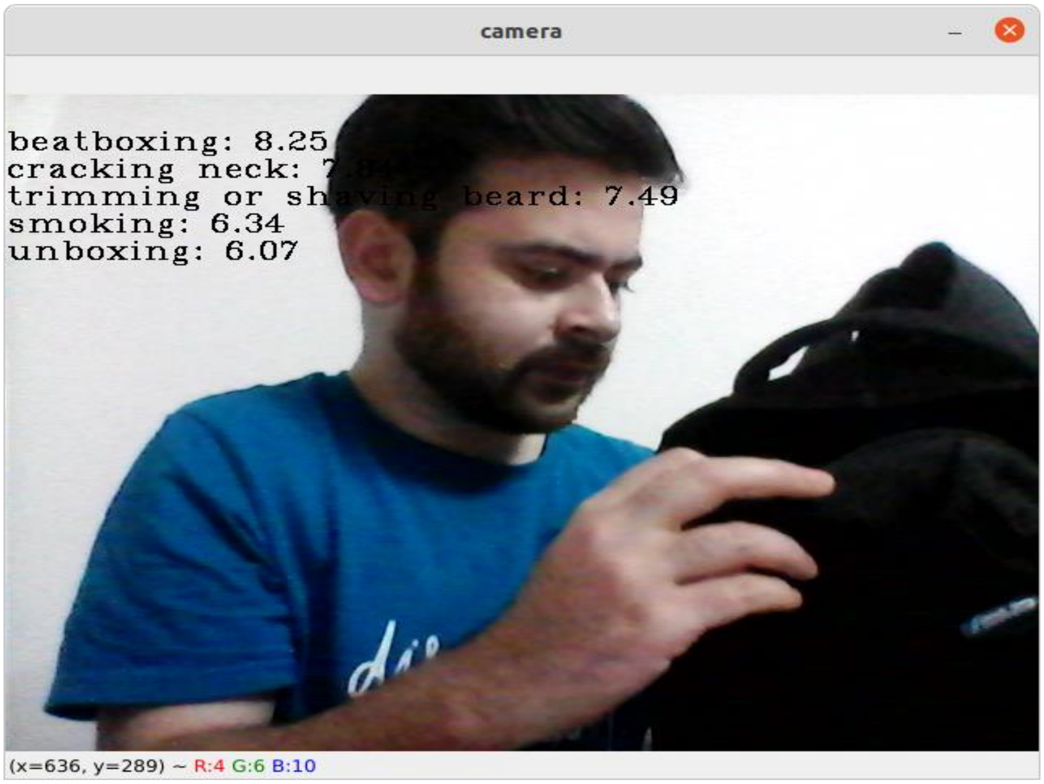
<u>Shaving Head / Getting a hair cut</u>

5.



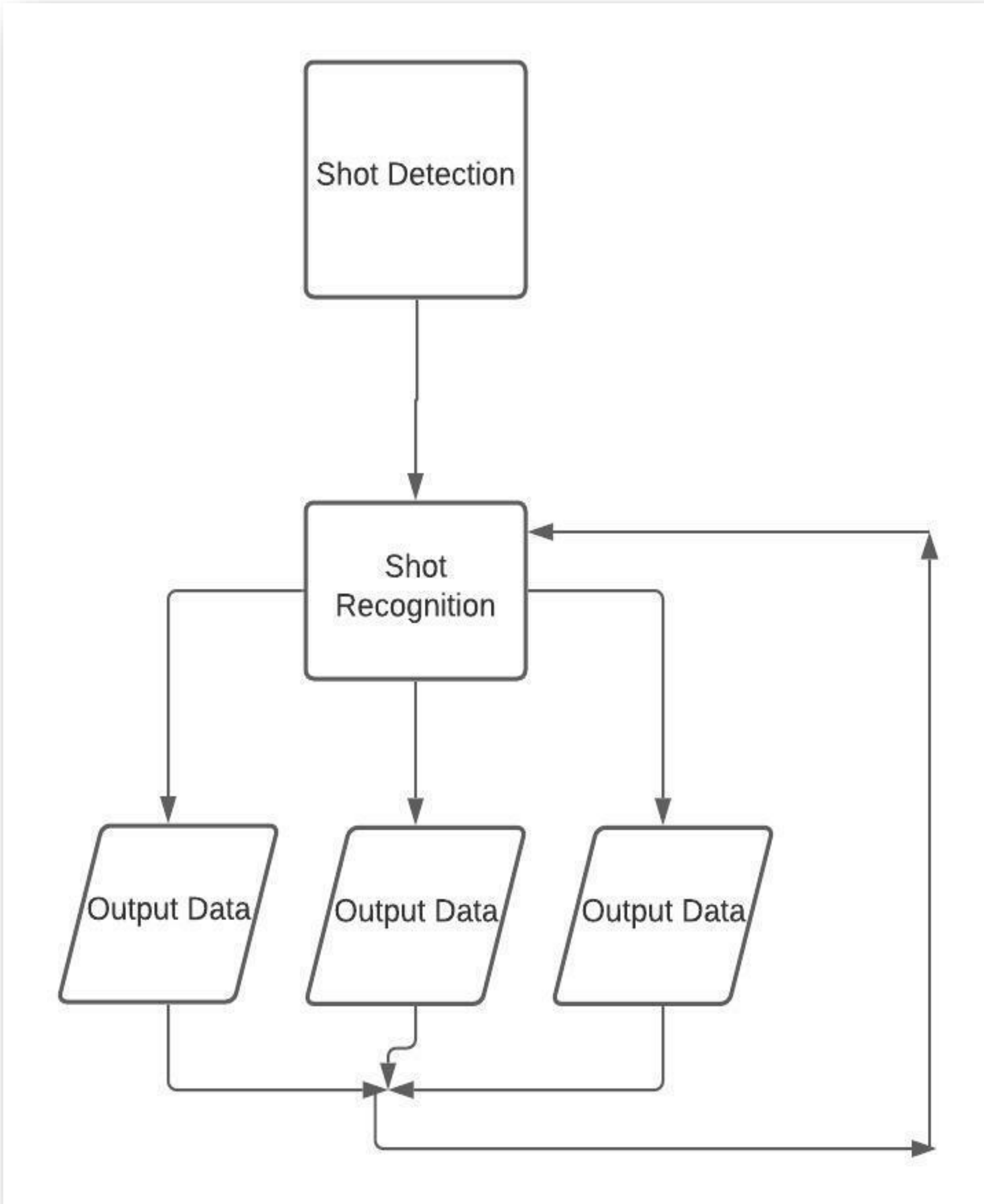
<u>Texting</u>

6.


camera
beatboxing: 8.25
beard: 7.49
smoking: 6.34
unboxing: 6.07
(x=636, y=289) ~ R:4 G:6 B:10

Unboxing

WORK FLOW DIAGRAM



CHAPTER - 4

Results and Discussion

We first evaluate on the early action recognition task, which aims to classify actions that have started, but not finished yet. In other words, the observed video is only partially completed, producing uncertainty. We use videolevel action labels to train the classification layer on zˆN (ct), for all time steps t. Tab. 1 and Tab. 2 show that, with everything else fixed, hyperbolic models learn significantly better representations than the Euclidean counterparts (up to 14% gain). The hyperbolic representation enjoys substantial compression efficiency, indicated by the 64 dimensional hyperbolic embedding outperforming the larger 256 dimensional Euclidean embedding (up to 5%). As indicated by both hierarchical accuracy metrics, when there is uncertainty, the hyperbolic representation will predict a more appropriate parent than Euclidean representations.

We next evaluate the representation on future action prediction, which aims to predict actions before they start given the past context. There is uncertainty because the next actions are not deterministic. A key advantage of hyperbolic representations is that the model will automatically decide to select the level of abstraction based on its estimate of the uncertainty. If the prediction is closer to the center of the Poincare ball, the model ´ lacks confidence and it predicts a parental node close to the root in the hierarchy. If the prediction is closer to the border of the Poincare ball, the model is more confident and ´ consequently predicts a more specific outcome. We fine-tune our model to learn to predict the class of the last clip of a video at each time step, for each of the three hierarchy levels in the FineGym dataset. We use cliplevel labels to train the classification layer on the model's prediction zˆN (ct). We select a threshold between hierarchy levels by giving each level the same probability of being selected: the predictions that have a radius in the smaller than the 33% percentile will select the more general level, the ones above the 66% percentile will select the more specific level, and the rest will select the middle level.3 Once the thresholds are set, we obtain both the predicted hierarchy level as well as the predicted class within that level. We report values for $t = N - 1$. For all three metrics, predicting a hierarchical representation substantially outperforms baselines by up to 14 points. The gains in both top-down and bottom-up hierarchical accuracy show that our model selects a better level of abstraction than the baselines in the presence of uncertainty.

The hyperbolic model also obtains better performance than the Euclidean model at the standard classification accuracy, which only evaluates the leaf node prediction. Since classification accuracy does not account for the hierarchy, this gain suggests hyperbolic representations help even when the future is certain. We hypothesize this is because the model is explicitly representing uncertainty, which stabilizes the training compared to the Euclidean baseline. Since our model represents its prediction of the uncertainty, we are able to visualize which videos are predictable. And hence this software efficiently using continual learning tries to predict future actions with high accuracy.

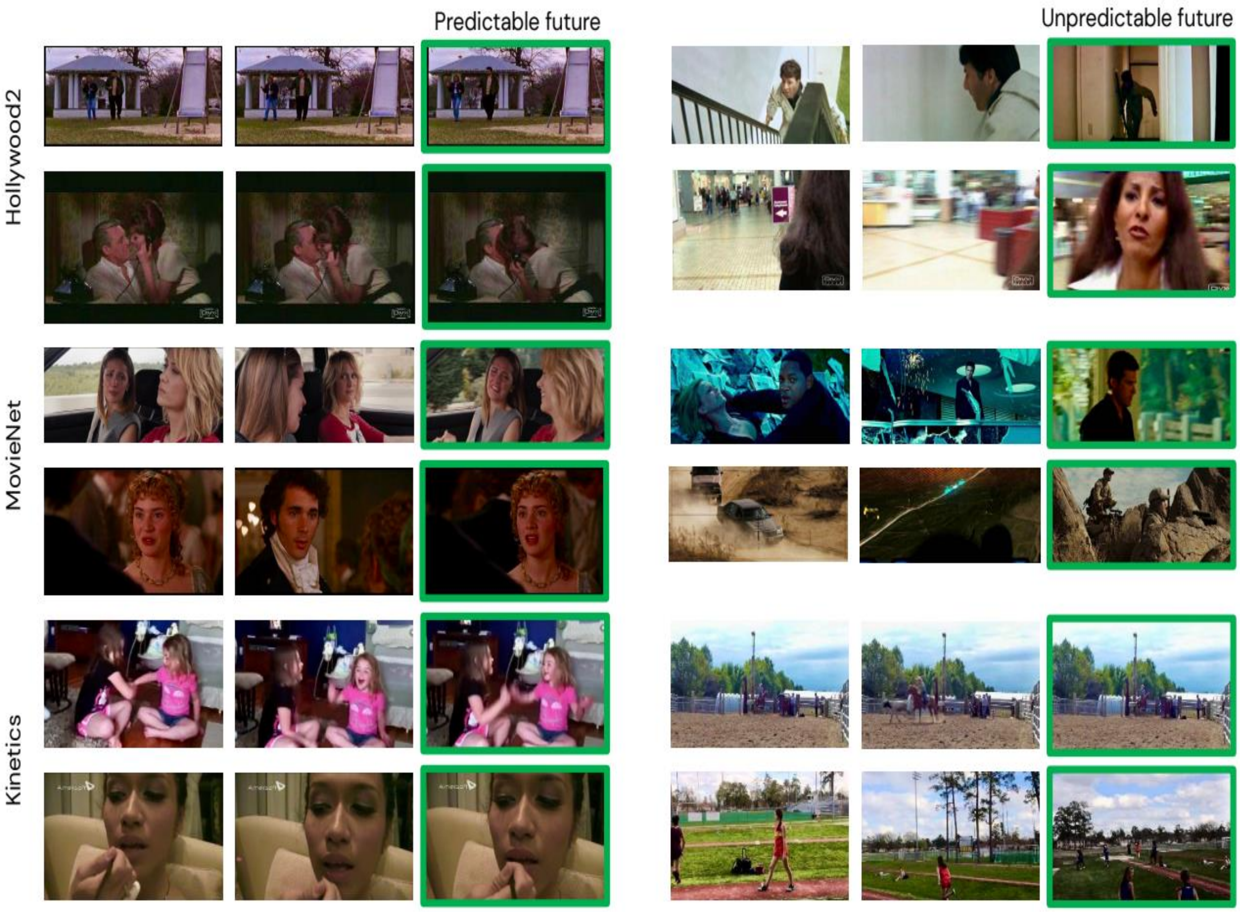


Figure 4. Examples with high predictability (left), and low predictability (right). The first two frames represent the content the model sees, and the frame in green represents the action the model has to predict. The high predictability examples are selected from above the 99 percentile and the low predictability examples are selected from below the 1 percentile, measured by the radius of the prediction.

CHAPTER – 5

CONCLUSION

While there is uncertainty in the future, parts of it are predictable. We have introduced a hyperbolic model for video prediction that represents uncertainty hierarchically. After learning from unlabeled video, experiments and visualizations show that a hierarchy automatically emerges in the representation, encoding the predictability of the future.

Refrences

1. Rodríguez-Moreno I., Martínez-Otzeta J.M., Sierra B., Rodriguez I., Jauregi E. Video Activity Recognition: State-of-the-Art. *Sensors.* 2019;**19**:3160. doi: 10.3390/s19143160.
2. Wren C.R., Azarbayejani A.J., Darrell T.J., Pentland A.P. *Integration Issues in Large Commercial Media Delivery Systems.* SPIE; Washington, DC, USA: 1996.
3. Elgammal A., Harwood D., Davis L. *Computer Vision—ECCV 2000.* Springer; Berlin, Germany: 2000. Non-parametric model for background subtraction; pp. 751–767.
4. Barnich O., Van Droogenbroeck M. ViBE: A powerful random technique to estimate the background in video sequences; Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing; Taipei, Taiwan. 19–24 April 2009; pp. 945–948.
5. McFarlane N.J.B., Schofield C.P. Segmentation and tracking of piglets in images. *Mach. Vis. Appl.* 1995;**8**:187–193. doi: 10.1007/BF01215814.
6. Lucas B.D., Kanadee T. Aniterative image registration technique with an application tostereovision; Proceedings of the Imaging Understanding Workshop; Pittsburgh, PA, USA. 24–28 August 1981; pp. 121–130.
7. Horn B.K., Schunck B.G. *Techniques and Applications of Image Understanding.* Volume 17. International Society for Optics and Photonics; Washington, DC, USA: 1981. Determining optical flow; pp. 185–203. Technical Symposium East.
8. Lowe D.G. Object recognition from local scale-invariant features; Proceedings of the Seventh IEEE International Conference on Computer Vision; Kerkyra, Greece. 20–27 September 1999;
9. Dalal N., Triggs B. Histograms of oriented gradients for human detection; Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05); San Diego, CA, USA. 20–25 June 2005;
10. Comaniciu D., Meer P. Mean shift analysis and applications; Proceedings of the Seventh IEEE International Conference on Computer Vision; Kerkyra, Greece. 20–27 September 1999;
11. Almeida, Jurandy, Ricardo da S. Torres, and Neucimar J. Leite. "Rapid video summarization on compressed video." *2010 IEEE International Symposium on Multimedia*. IEEE, 2010.
12. Elharrouss, Omar, et al. "A combined multiple action recognition and summarization for surveillance video sequences." *Applied Intelligence* 51.2 (2021): 690-712.
13. Hadsell, Raia, et al. "Embracing change: Continual learning in deep neural networks." *Trends in cognitive sciences* (2020).
14. Parisi, German I., et al. "Continual lifelong learning with neural networks: A review." *Neural Networks* 113 (2019): 54-71.

15. M. McCloskey and N. J. Cohen, “Catastrophic interference in connectionist networks: The sequential learning problem,” Psychology of learning and motivation, vol. 24, pp. 109–165, 1989.
16. Aimone, J.B., Wiles, J., Gage, F.H.: Computational influence of adult neurogenesis onmemory encoding. Neuron 61, 187–202 (2009)
17. Borji, A., Izadi, S., Itti, L.: iLab-20M: A Large-Scale Controlled Object Dataset to In-vestigate Deep Learning. In: International Conference of Computer Vision and PatternRecogniton (CVPR), pp. 2221–2230 (2016)
18. Parisi, G. I. (2020). Human action recognition and assessment via deep neural network self-organization. In *Modelling Human Motion* (pp. 187-211). Springer, Cham.
19. An, Jiahui, et al. "Summary of continuous action recognition." *Journal of Physics: Conference Series*. Vol. 1607. No. 1. IOP Publishing, 2020.
20. Li, Zhizhong, and Derek Hoiem. "Learning without forgetting." *IEEE transactions on pattern analysis and machine intelligence* 40.12 (2017): 2935-2947.
21. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014.
22. G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” in NIPS Workshop, 2014.
23. H. Jung, J. Ju, M. Jung, and J. Kim, “Less-forgetting learning in deep neural networks,” arXiv preprint arXiv:1607.00122, 2016.
24. Li, Zhizhong, and Derek Hoiem. "Learning without forgetting." *IEEE transactions on pattern analysis and machine intelligence* 40.12 (2017): 2935-2947.
25. R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014.
26. G. Hinton, O. Vinyals, and J. Dean, “Distilling the knowledge in a neural network,” in NIPS Workshop, 2014.
27. H. Jung, J. Ju, M. Jung, and J. Kim, “Less-forgetting learning in deep neural networks,” arXiv preprint arXiv:1607.00122, 2016.