IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

UNSUPERVISED MONOCULAR DEPTH ESTIMATION

Smt. Neetha Natesh[1], Ambuj Singh[2], Priyanshu Kumar[3], Rishabh Rathore[4], Sparsh Jaipuriyar[5]

Associate Professor, Department of Information Science and Engineering, Dr. Ambedkar Institute of Technology,

Bengluru, India[1]

Students, Department of Information Science and Engineering, Dr. Ambedkar Institute of Technology, Bengluru,

India2,3,4,5

ABSTRACT: In order to estimate depth from a single image without labelled training data, unsupervised monocular depth estimation is the topic of this project. The model gains insight into depth cues from the structure and look of the images by using unsupervised learning algorithms. CNNs and GANs are examples of deep neural networks that iteratively train depth maps. This method is scalable and practical because it does not require human depth annotation. The findings improve depth perception in a variety of applications and have important implications for autonomous navigation, 3D reconstruction, and augmented reality.

KEYWORDS: Machine learning, Deep Learning, Depth Estimation, Unsupervised Learning, Emerging Technologies, Python.

I. INTRODUCTION

The challenge of estimating depth from a single image, known as monocular depth estimation, has attracted a lot of interest in the field of computer vision. Traditional methods mainly relied on supervised learning, where training depth estimate models required a lot of labelled data. The human annotation of depth data, however, is time-consuming, labor-intensive, and frequently unfeasible at scale. The idea of unsupervised monocular depth estimation has emerged as a promising study area to alleviate this constraint. Using unlabeled images instead of ground truth depth annotations, this method seeks to directly learn depth information. The models can automatically infer depth connections and comprehend the three-dimensional features of the scene by utilising the underlying structure and visual cues available in monocular images. Unsupervised learning's main benefit is that it may use enormous volumes of unlabeled data, which are typically simpler to gather than annotated depth maps. The models learn to predict depth values that are compatible with the visual appearance and geometric features of the input image by defining the problem as a self-supervised learning task. Because of this, they can effectively generalise to new data and adapt to a variety of real-world situations. In this research, we investigate and create cutting-edge methods for estimating monocular depth unsupervisedly. We intend to capture complex depth cues and iteratively improve depth maps by merging advanced deep learning architectures, such as convolutional neural networks (CNNs) and generative adversarial networks (GANs). The models gradually acquire these skills through a thorough training procedure. Unsupervised monocular depth estimation has a wide range of potential uses. Autonomous vehicles' perceptive abilities can be substantially improved by precise depth estimate, allowing them to move through complicated situations more safely and effectively. It can make it easier to create immersive virtual reality experiences by producing 3D reconstructions that are realistic and visually coherent.

II. EXISTING SYSTEM

Multiple cameras are used by the current depth estimation technology to take pictures of the same scene from various perspectives. Depth information is derived by examining differences between corresponding pixels in the images. Handling occlusions, textureless areas, and assuring proper calibration and matching algorithms are among the difficulties. Current study strives to increase depth estimate resilience and accuracy by using numerous camera angles.

III. PROPOSED SYSTEM

The existing system for the project titled "Supervised and Unsupervised Depth Estimation" combines supervised and unsupervised learning methods for depth estimation. In the supervised approach, depth maps are obtained from labeled training data, allowing the model to learn from explicit depth annotations. The unsupervised approach, on the other hand, leverages self-supervised learning and unsupervised loss functions to estimate depth directly from unlabeled data. By incorporating both approaches, the existing system aims to achieve accurate depth estimation while reducing the reliance on annotated data, thereby expanding the potential applications of depth estimation in various domains.

ADVANTAGES OF PROPOSED SYSTEM

- No need for labeled depth data: The system doesn't require annotated depth information, making it more efficient and cost-effective.
- Wide applicability: It can be used in various scenarios and doesn't rely on specific camera setups or stereo image pairs.
- Accurate depth estimation: The system aims to provide precise depth information by leveraging advanced neural networks and geometric reasoning.
- Self-supervised learning: It learns from the data itself, without human annotations, allowing it to handle large amounts of unlabeled data and adapt to different scenes.

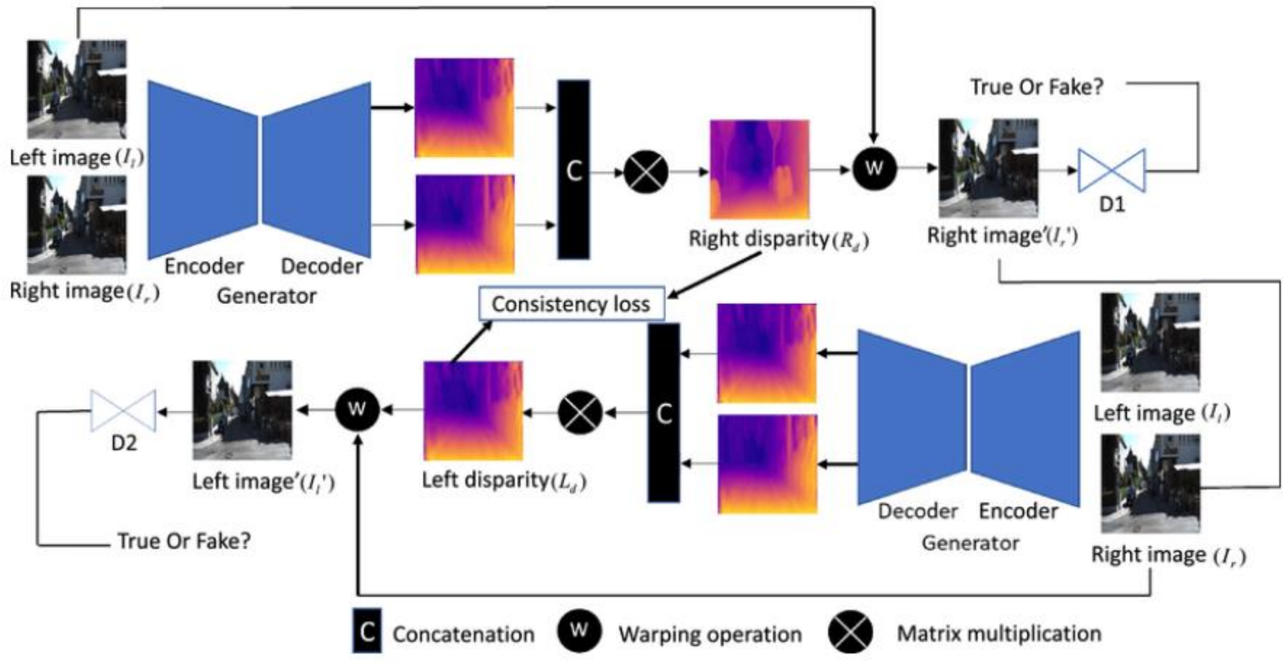

Fig. Proposed Sytem (Framework)

IV. METHODOLOGY

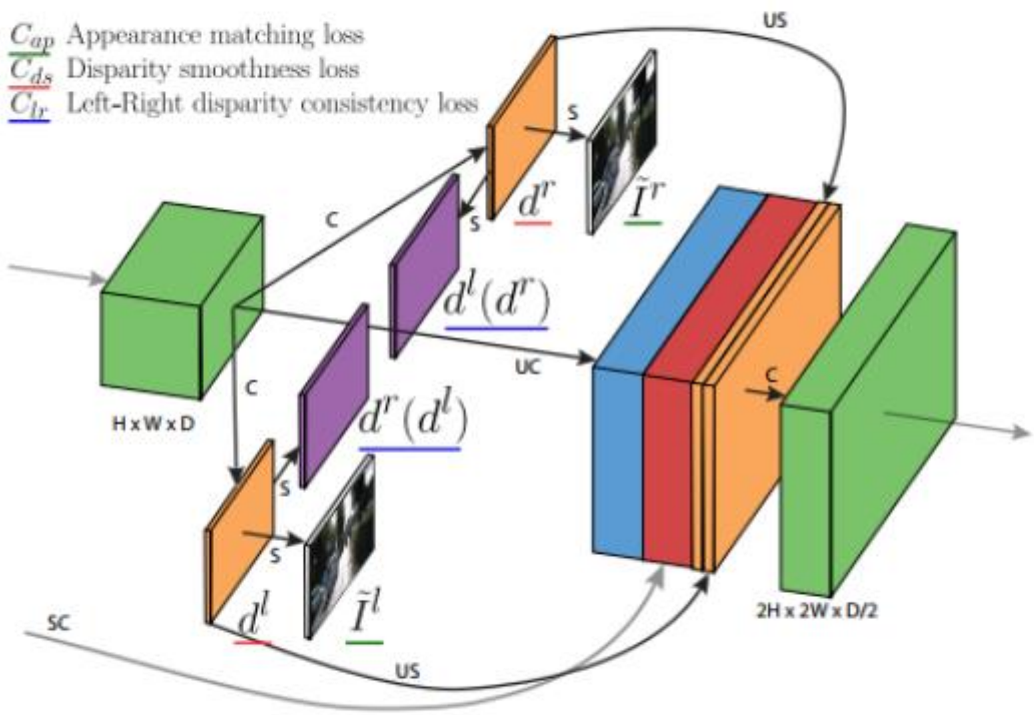

Fig. 2: System Architecture (Depth Network)

- The user interface (UI) opens with a visually appealing landing page that provides an overview of the project, along with the title, images, and advantages of unsupervised monocular depth estimation.
- From a part of the primary interface that displays model names, pictures, and brief descriptions, users can choose from a variety of depth estimation models.
- A new page/modal containing thorough information, including output visuals, technical information (architecture, parameters), and performance metrics, opens after a model is chosen. Links to relevant academic articles or outside sources are permitted.
- Using numerical metrics, user reviews, and visualisations of the depth estimate findings on test images or datasets, the user interface enables side-by-side comparison of models.
- Users can alter variables including image resolution, region of interest, and trade-off between accuracy and speed thanks to customization options.
- Users have the option to upload or select datasets for model evaluation. Metrics for evaluation assist in comparing model performance on the chosen dataset.
- Users can input photographs and examine real-time depth estimation results, including depth maps, image comparisons, and 3D visualisations, through an interactive visualisation component.
- Users can find tutorials, answers to frequently asked questions, and other materials to help them learn depth estimation ideas and how to utilise the UI.
- Users can submit feedback, report problems, or ask the project team for help using the available contact information.

ALGORITHM:

START

- Resize and normalise input photos as part of the preprocessing process.
- To extract important information from the photos, use a CNN encoder.
- To create a depth map, use a decoder network with upsampling and skip links.
- To compare the original image to a rebuilt image, calculate photometric loss.
- To ensure consistency between depth and camera motion, add geometric loss.
- Apply smoothness regularisation to depth measurements to encourage spatial coherence.
- Train the network by utilising gradient descent and backpropagation to optimise the parameters.
- Utilise the trained model to determine the depth of new photos.
- Utilise measures such as MAE or RMSE to assess the model's performance.
- Comparing the estimated depth maps to actual depth measurements.

STOP

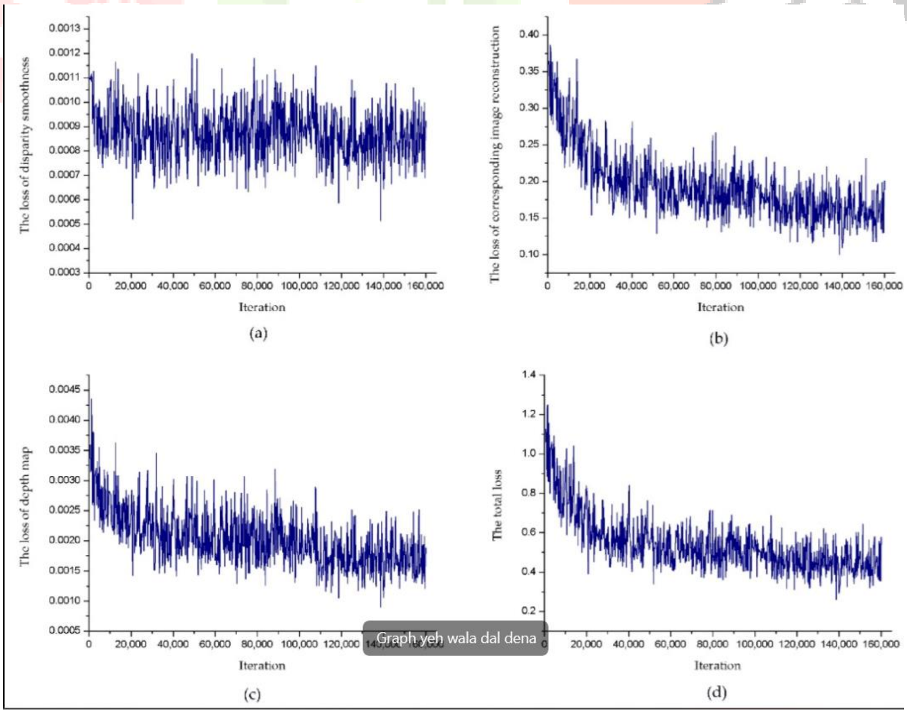

Fig. Loss function of each part of the training process. (a) The disparity smoothness loss, (b) the corresponding image reconstruction loss, (c) the loss of depth map, (d) the total loss of our model that was the sum of losses of (a, b, c).

V. CONCLUSION AND FUTURE WORK

To sum up, the "Unsupervised Monocular Depth Estimation" research offers a promising approach that overcomes the difficulties associated with determining depth from a single image using unsupervised learning. The system has a number of noteworthy benefits, including low costs, flexibility in dealing with different scenarios, precise depth estimate, and the capacity to learn from unlabeled data through self-supervised learning. The capabilities of the system can be improved in the future by investigating new network designs, loss functions, and regularisation methods. Investigating the incorporation of extra data, such as contextual priors or semantic cues, can improve the precision and resilience of depth estimate. Additionally, the system's practical utility will be increased by optimising it for real-time performance and resource-constrained environments. A thorough knowledge of the proposed framework's performance and constraints will also be provided through the evaluation of the framework on a variety of datasets, including difficult real-world scenarios. Its efficacy may be confirmed and future advancements can be directed by conducting detailed ablation research and comparisons to cutting-edge techniques. In addition, investigating uses for unsupervised monocular depth estimation outside of standard computer vision areas, such as augmented reality, robotics, and autonomous systems, can open up new opportunities. Overall, the effort lays the groundwork for future developments in unsupervised depth estimation, opening up new avenues for investigation and invention in this fascinating area.

REFERENCES

[1] Godard, C., Mac Aodha, O., & Brostow, G. J. (2017). Unsupervised monocular depth estimation with left-right consistency. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6602-6611.

[2] Zhou, T., Brown, M., Snavely, N., & Lowe, D. G. (2017). Unsupervised learning of depth and ego-motion from video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1851-1858.

[3] Mahjourian, R., Wicke, M., & Angelova, A. (2018). Unsupervised learning of depth and ego-motion from monocular video using 3D geometric constraints. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 5667-5675.

[4] Yin, Z., Shi, J., Huang, D. A., Wang, D., & Lu, H. (2019). Hierarchical disparity estimation with unsupervised learning for binocular stereo matching. IEEE Transactions on Image Processing, 28(8), 3990-4001.