WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

An AI Driven Approach for Smoke and Haze Elimination Algorithm

Akash R, Raja Abilash J, Gowthaman R

Rathinam College of Arts and Science

DOI: https://doi.org/10.55248/gengpi.5.0524.1423

ABSTRACT

The target of video dehazing is to reestablish fog free edges with amazing perceivability and difference in indoor fire mishaps through CCTV for the acknowledgment of the people who stuck under these fire risks and salvage them. Multi-scale Dehazing Combination (MDF) model with Profound Learning techniques like DCP (Dark Channel Prior) and TMAP (Transmission map) used to recuperate the corrupted cloudiness recordings by reproducing the video in high perceivability for tracking down the people in question. The model pre-processes the cloudy video into outlines. The foggy casings are then taken care of into the DCP and TMAP cycle to identify the smoke and barometrical light assessment from each edge. DMF model will encode and interpret each casing to manage debased recordings and Multi combination process supplies free data from one edge to another. A recreation module intended to remake high goal of cloudiness free video outlines. The result will help the salvage group to perceive individual through the web interface. This model was assessed on a dataset of 60000 murkiness outlines/pictures. The model accomplished with an exactness of 98% on testing with given input murkiness video which assists fire with safeguarding group to distinguish individuals who stuck under the fire situations. This could work on the precise identification of casualties in the indoor fire mishaps.

Keywords: Video dehazing, multiscale attention, dilated convolution

1. INTRODUCTION

In murky precipitation conditions, videos caught by imaging predisposition are for the most part very discouraged. Therefore, tape dehazing is of extraordinary importance and wide in out-of-entryway reconnaissance and autonomous driving, and it has drawn in broad consideration. The improvement cycle of the foggy picture is for the most part portrayed as follows I(x) = J(x) t(x) A(1 − t(x))(1) where I(x) is the gathered murky bad quality picture, J(x) is the cloudiness free picture, t(x) is the conveyance outline, An is the worldwide environmental light, and x is the pixel organizes. A murky tape can be seen as a constant grouping of cloudy pictures. in any case, there are a few equals among tape and picture dehazing. tape dehazing eliminates the cloudiness of a solitary edge and thinks about the equal data in progressive casings. Contrasted with picture dehazing, the principal challenge of tape dehazing is the means by which to straightforwardly adjust commensurate edges to reference outlines and remake a cloudiness free picture. Given the huge headway made in picture dehazing inside PC vision, most extreme experimenters have attempted to stretch out this accomplishment to tape dehazing by performing picture transformation on an edge by-outline base. in any case, this approach neglects to think about the proportional data between progressive edges, prompting extreme remnants. While a few styles for tape dehazing have surfaced as of late, none have straightforwardly adjusted commensurate edges to the reference outline. additionally, certain tape dehazing styles can not reproduce the dehazed tape successfully as a result of the great goal of true tape's dehazing(REVIDE) dataset. This study proposes a Multi-scale Dehazing Combination(MDF) for clear dimness free videos to address the previously mentioned difficulties. MDF utilizes an encoder-decoder structure grounded on multiscale thoughtfulness regarding recognize and intertwine highlights at various circumstances during the point birth stage. The point extractor utilizes Res2Net(1) as the encoder, and the decoder utilizes the proposed point emulsion subcaste(FFL) grounded on channel and pixel consideration. This builds the open field of each subcaste and gives firmness for managing dampened outlines. The MDF summations equal data between progressive edges in the multi-outline emulsion stage utilizing a planned pixel consideration directed emulsion module(AMF) to adjust numerous casings. In the remaking phase of verifiable edges, MDF utilizes projected swelled difficulties with various development rates to expand the open field and better reproduce high-goal dormant casings. The far reaching preliminaries led on the REVIDE dataset exhibit that MDF performs well than different styles with respect to tape dehazing execution. In outline, the gifts of this study are as per the following • We propose another Multi-scale Dehazing Combination(MDF), and extensive preliminaries are directed on REVIDE datasets to exhibit the adequacy of MDF.

• A point extractor grounded on multiscale consideration is intended to prize and wire highlights at various circumstances. This gives resoluteness for managing unsettled videos.

• A pixel AMF lays out cross-outline correspondences and summations complementary data from one casing to another.

• A reproduction module grounded on distended expanded difficulties is intended to recreate high-goal inactive approaches more and give the likelihood to remake videos of various decisions .

1.1 MOTIVATION

The ongoing scene of learning-based strategies transcendently takes care of picture handling, neglecting the basic domain of video dehazing. This hole represents a test in consistently extricating and amalgamating spatial and transient qualities across progressive edges while taking a stab at a genuine casing rebuilding. Enter the Multi-scale Dehazing Combination (MDF), a historic arrangement that expects to restore clearness inside dehazed recordings. MDF pioneers a multiscale consideration based highlight extraction instrument, digging into different levels for far reaching highlight obtaining. Its creative methodology incorporates a pixel consideration directed multi-outline combination module, decisively winding around together worldly components and orchestrating data from connecting outlines. Besides, with a remaking module utilizing flowed widened convolutions, MDF endeavors to recreate idle edges carefully. The exact approval through broad preliminaries on true video dehazing datasets grandstands the unparalleled ability of MDF. Important are its wonderful exhibition measurements, bragging PSNR and SSIM scores 23.93dB and 0.7732 separately, setting another norm in video dehazing techniques.

1.2 CONTRIBUTION

This examination makes a huge commitment by tending to the underexplored space of video dehazing inside learning-based strategies, which have basically focused on picture handling. The presentation of the Multi-scale Dehazing Combination (MDF) marks a critical progression in this field, offering a hearty answer for reestablishing lucidity and eliminating fog from recordings. The center commitment lies in MDF's creative philosophy, utilizing a multiscale consideration based highlight extractor that exhaustively catches highlights across different levels. Moreover, the coordination of a pixel consideration directed multi-outline combination module empowers compelling fleeting displaying, empowering consistent mix of fundamental data from adjoining outlines. Furthermore, the consideration of a recreation module using flowed enlarged convolutions essentially upgrades the accuracy in remaking dormant edges. Through thorough trial and error on certifiable video dehazing datasets, MDF grandstands its predominance over existing techniques, as confirmed by significantly better execution measurements, including PSNR and SSIM scores coming to 23.93dB and 0.7732, separately. This commitment lays out MDF as a spearheading structure in video dehazing, setting new benchmarks for viability and execution inside this specific space.

RELATED WORK

A. IMAGE DEHAZING

Single picture dehazing ways are basically arranged into physical-grounded and profound learning styles.

1. PHYSICAL-BASED STYLES

In the beginning phases of the single-picture dehazing, actual models and bright picture past information or hypotheticals are by and large utilized. For outline, introduced the dull channel past dehazing framework, which utilized the past information on dim essential tones to advance transmission graphs. chipped away at the past procedure for dehazing of variety lessening, which breaks down various murky pictures and notices the consideration of murkiness changes with field profundity. a calculation grounded on non-unique priors to acquire clear and dimness free pictures. To acquire the reestablished picture, utilized middle sifting to assess the residue and cloudiness consideration.

2.PROFOUND LEARNING STYLES

This framework, otherwise called the information driven framework, includes utilizing brain organizations to become familiar with the planning of a relationship between a cloudy picture and its adjusting cloudiness free picture for picture dehazing. For case, Cai etal.(1)pro-presented Dehaze Net, which utilized a convolutional brain network(CNN) to encode transmission diagrams and worldwide climatic light. Li etal.(2) proposed AODNet, which meant to appraise the worldwide air light and transmission outline contemporaneously. Zhang etal.(3) To upgrade picture dehazing, the" Thickly Related Total Dehazing Organization" was introduced utilizing transmission diagram and worldwide barometrical light prior to acquiring the dehazed picture. Shao etal.(4) proposed the versatile circle outline work for picture dehazing. Other remarkable methodologies incorporate GFN(5), a limit grounded emulsion network that gauges the comparing weight of the changed over picture through the encoder-decoder organization and breakers them to acquire the last dehazed results. GCANet(6) utilizes expanded intricacy to bar network remnants. FFA-Net(7) is a start to finish point emulsion consideration network that consolidates channel and pixel consideration instruments to give model resoluteness. TDN(8) is another model from coarse to relinquishment that learns the planning among dim and murkiness free pictures through programmed fog consistency acknowledgment. MSBDN-DFF(9) planned a SOS-upgraded module to finish the job of dehazing. GridDehazeNet(10) is a teachable start to finish CNN that devices consideration grounded multiscale assessment in the spine module to whitewash the running back issue in conventional multiscale styles. Sun et al(11) proposed a versatile dehazing Siamese organization to acquire dehazed pictures with clear edges. TBNet(12) is a three-branch dehazing network grounded on subspace projection, which accomplishes great dehazing results. As referenced previously, the framework has accomplished magnificent dehazing execution in picture dehazing; still, it can not successfully complete the tape dehazing task since it doesn't consider the fleeting connection between tape outlines.

B. MULTI-CASING ARRANGEMENT

Of late, experimenters have proposed a few styles for adjusting commensurate casings to reference outlines. These styles incorporate Markov erratic fields (MRF), optic inflow, deformable intricacies, and brain organizations. Xiang etal.(12) first utilized optic inflow assessment to investigate worldly

data and total adjoining information. additionally, they proposed an encoder - decoder network that investigates clear elements from models to direct better picture rebuilding. Wang etal.(12) proposed utilizing stirred up deformable difficulties with kill evaluations to arrange coextensive edges to reference outlines in point space. additionally, they utilized spatial and transient consideration modules to combine these highlights. Zhong etal.(12) first utilized an irregular brain organization and a worldwide spatiotemporal consideration module to adjust commensurate casings to the reference outlines. likewise, they utilized a reproduction organization to recover the adjusted reference outline. Caballero etal.(13) proposed utilizing spatio-transient sub-pixel intricacy and normal mix remuneration to investigate fleeting overt repetitiveness and enhance reproduction delicacy really. Zhang etal.(13) utilized three-layered (3D) inconveniences to coincidently reuse spatial and fleeting data by lazily intertwining different progressive tape outlines. The forenamed system, still powerful in yielding noteworthy issues, experiences a trouble in light of the prominent ascent in computing intricacy achieved by the use of 3D complexity. Put further essentially, the technique works however is computationally savage since it needs a huge quantum of handling power. Su etal.(14) consolidates optic inflow and super-pixel matching to accomplish fine-granulated arrangement of scene content at pixel position and semantic position, and uses optic inflow to preprocess worldwide edge arrangement. likewise, customary styles can not straightforwardly gauge boundaries comparable as mix data and kill at times (for representation, enormous mix and thick murkiness), prompting a drop in the presentation of tape dehazing.

C.VIDEO DEHAZING

Multiple ways have been proposed for tape dehazing. For case, Li etal.(15) introduced a framework to contemporaneously encode the scene profundity and enhance clear idle edges from foggy tape successions. Cai etal.(16) raised a model grounded on Markov inconsistent field(MRF) utilizing force values, which settle the spatial and fleeting data of three commensurate edges and upgrades the MRF responsibility capability to recuperate the reference outline. Wang etal.(16) proposed a dispersed casing for ongoing tape dehaz ing adjusting of three fundamental variables climatic light assessor, transmission outline assessor and cloudiness free picture maker. EDVNet(17) utilized three convolutional layers to send three commensurate edges into point space. These highlights were jutted and moved into the dehazing organization to accomplish tape dehazing. Ren etal.(18) proposed a profound proficiency result for tape dehazing, where the organization gauges the transmission diagram by gathering data across outlines. This framework utilizes the worldwide semantic past as contribution to homogenize the transmission diagram and furthermore reestablishes the murkiness free casing through the barometrical dissipating model. Li etal.(19) proposed a tape inpainting framework to eliminate debasements and naturally prompt clean videos. This framework figured on the consideration outline and finished spills over to direct the information outline recuperation process, where clean pixels are accomplished from coextensive edges utilizing the proposed irregular design. Zhang etal.(19) proposed a certainty directed and improved deformable organization to resolve the issue of taking advantage of worldly overt repetitiveness among the dim casings in accomplishing tape dehazing. A high level framework for adjusting and reestablishing tape dehazing. In the arrangement cycle, the framework gradationally adjusts progressive commensurate casings without utilizing optic inflow assessment, and it utilizes a refinement network in the recuperation interaction to enhance the dehazing execution. It likewise introduced a relevant unfriendly proficiency network with reproducing point sharing for tape dehazing. The organization utilizes a multiscale swelled lingering module to catch worldwide and rich context oriented data and coordinates a channel consideration block with the as of late introduced guide conglomeration module toward recalibrate the loads of the showing up point diagrams. The styles referenced over have accomplished brilliant dehazing execution, however the outcomes are often off-base when applied to high-goal videos.

3. METHODOLOGY

3.1 PROBLEM STATEMENT

The test of proficiently saving people caught in indoor or outside fire dangers is intensified by restricted perceivability brought about by smoke and dimness. The requirement for an artificial intelligence ML based keen de-smoking/de-right of passage calculation emerges to resolve this basic issue by handling ongoing recordings got from live feed CCTV sources. The proposed arrangement, using the MDF model, plans to successfully de-smoke and de-fog the information video by utilizing pre-prepared highlights. The essential goal is to upgrade perceivability inside these recordings, permitting salvage groups to more readily explore and distinguish people caught in these dangerous situations. By giving a more clear, de-hazed video yield, the calculation tries to offer essential help by means of a web interface, supporting salvage groups in perceiving and finding people in trouble in the midst of the difficult circumstances presented by fire occurrences.

3.2 PROPOSED STRATEGY

This segment presents the overall armature of the proposed MDF was first presented for tape dehazing, trailed by the nitty gritty plan of each piece of the organization and effective for better dehazing performance.

DARK CHANNEL PRIOR

Earlier The Dark Channel Prior past is grounded on the accompanying perception on out-of-entryway dimness free pictures The low power in obscurity channel is significantly because of three factors

a) murk,e.g., the murk of transports , structures, and within windows in cityscape pictures, or the murk of leaves, trees, and gems in geology pictures;

b) different items or shells,e.g., any article with low reflectance in any variety channel(for representation, green grass/tree/manufacturing plant, red or unheroic bloom/support, and blue water face) will influence in low qualities in obscurity channel;

c) dim articles or shells,e.g., dim tree caddies and landmarks. As the normal out-of-entryway pictures are by and large different and brimming with murk, the dark channels of these pictures are truly dim! To verify how great the dark channel past is, we gather an out-of-entryway picture set fromFlickr.com and a few other picture chase machines utilizing 150 most famous markers commented on by the Flickr junkies. Since dimness by and large happens in out-of-entryway geology and cityscape scenes, we physically select the cloudiness free topography and cityscape bones from the informational index. possibly, we just focus on day pictures. Among them, we randomly choose 5,000 pictures and physically cut out the sky districts. The pictures are resized so the beyond reach and level is 500 pixels and their dim channels are figured utilizing a fix size 15 × 15. This shows a few out-of-entryway pictures and the comparing dim channels. Our dark channel past is not entirely propelled by the notable dim item derivation design widely utilized in multispectral remote seeing frameworks. The spatially homogeneous murkiness is eliminated by lessening a steady worth comparing to the most obscure item in the scene. We sum up this thought and propose a new past for normal picture dehazing.

TRANSMISSION MAPS

Transmission maps play a critical role in understanding the degree of light attenuation caused by smoke and haze. These maps represent the ratio between the observed degraded image and the underlying scene that's obstructed by smoke particles. In essence, they depict the level of transparency or haziness present in different regions of the video frames. By estimating these transmission maps, algorithms can discern areas with varying degrees of smoke density, enabling a more targeted approach to haze removal. Leveraging these maps allows for selective enhancement, emphasizing clearer regions while attenuating the impact of smoke and haze in more heavily affected areas. Transmission maps serve as indispensable guides in the dehazing process, aiding in the restoration of visual clarity and facilitating improved perception for rescue operations in fire incidents.

FEATURE EXTRACTION

The proposed point extractor presents an encoder-decoder structure. The construction involves Res2Net as the encoder in the point birth phase of the MDF due to its ability to address multiscale highlights at the grainy position and increment the open field of each subcaste. In particular, the front facing part of Res2Net with multiple times down-cut was upheld, and totally associated layers were deserted. For the decoder part, the planned FFL was utilized as the base subcaste of various scales, and the highlights of the relating scales in the encoder were linked to value the elements of progressive edges straightforwardly presents the FFL structure adjusting of convolutional layers, channel consideration, and pixel consideration. As indicated by channel and pixel consideration treats various highlights and pixel districts contrarily. This gives new resoluteness while taking care of various sorts of data and expands the portrayal capacity of CNN shows the designs of channel and pixel consideration.

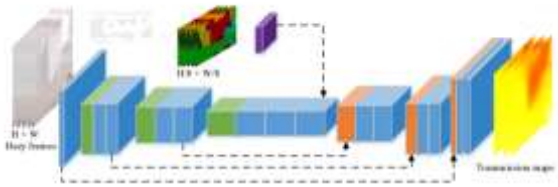
MULTI-CASING COMBINATION

In this segment, the proposed AMF summations complementary data between progressive casings to adjust different edges. This shows the construction of the proposed AMF, which comprises of 1 × 1 confusion and pixel consideration. In the first place, the issue point outline of the point extractor was subjected to a connect activity. Simultaneously, pixel-position consideration was applied to the direct outline of the inactive edge toward gain the consideration direction diagram. likewise, the linked point diagram was increased pixel-by-pixel alongside the consideration directed outline of the inactive edge to acquire equal data between the dormant and different casings. Ultimately, the accomplished proportional data graph was added pixel-by-pixel to the direct diagram of the inactive casing toward accomplish the fleeting displaying between various edges presents the construction of the pixel consideration.

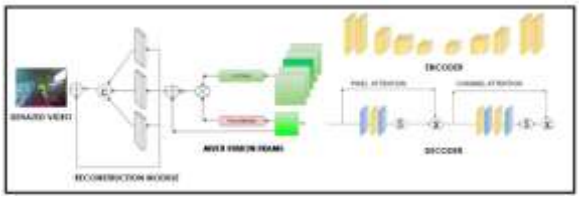
RECONSTRUCTIVE MODULE

The proposed MDF last stage reproduces the melded highlights utilizing a remaking module to deliver a fog free outcome Dt . Given the moderately high video goal (2708× 1800) in the REVIDE dataset, utilizing remaking modules with huge responsive fields is ideal. Subsequently, a leftover construction containing expanded convolutions with expansion paces of 1, 2, and 3 were utilized as the recreation module of MDF. In Figure 3, the design of the reproduction module is introduced.

3.3 PROPOSED FRAMEWORK



[DCP-TMAP]



[MDF MODEL]

3.4 DATASET

Haze World Dataset:

Since the ongoing tape dehazing datasets are considerably gathered in inward scenes, we develop a huge scope manufactured out-of-entryway tape dehazing dataset named Murkiness World, with representation outlines displayed in Information assortment. The first videos of Dimness World are from six being datasets, i.e., Cityscapes(10), DDAD(16), UA-DETRAC(54), Vis Robot(62), DAVIS(41), and REDS(39), acting in 1,271 cloudiness free videos. We utilize the environmental dispersing model to combine murky videos. The strong tape profundity assessment framework is utilized to acquire transiently amicable profundity graphs. This concentrate likewise utilized the REVIDE dataset and Custom Fire Mishap videos to prepare the proposed model. The REVIDE dataset contains inward scenes with high-depiction, containing 2031 dyads of unsettled and clear casings. Among these dyads, 1747 were utilized as preparing sets, and 284 were utilized as test sets. additionally, the goal of each casing in the dataset was 2708 × 1800 pixels. The Custom Dataset with Fire Pictures and videos from CCTV cameras are prepared with the proposed model.

RESULT AND DISCUSSION

QUALITATIVE ASSESSMENT

The visual examination of the proposed MDF framework with best in class dehazing styles utilizing the REVIDE dataset. In the subjective correlation show that the outcomes achieved utilizing the DCP and Plunge styles are more obscure in splendor and variety. The consequences of AODNet and VDN styles can not completely eliminate the impact of murkiness. additionally, the aftereffects of VOLUME 11, 2023 94483C. Han et al. tape Dehazing Organization Grounded on Multiscale Consideration. TABLE 1. Normal PSNR and SSIM consequences of various styles utilizing the REVIDE dataset. The FFANet and EDVR styles experience the ill effects of huge remnants. Albeit the Two-Branch framework accomplished improved brings about notwithstanding the impact of cloudiness, it loses variety data and certain subtleties. Contrasted with different styles, the MDF accomplishes the beautiful dehazing execution, close to the ground verity picture. This shows that the proposed pixel AMF better summations equal data between numerous edges, while the proposed projected swelled intricacy more reproduces inactive casings.

Method	PSNR	SSIM
DCP	10.04	0.8375
AODNet	15.01	0.7174
FFANet	21.01	0.9617
Two-Branch	23.49	0.9129

Method	PSNR	SSIM
DCP	14.01	0.8101
VDN	15.24	0.7642
EDVR	20.65	0.8179
MDF	28.93	0.7732

[TABLE 1]

QUANTITATIVE ASSESSMENT

For measurement, two well known measurements: PSNR and SSIM, were utilized for the calculations, and Table 1 presents the trial results. The outcomes acquired from the MDF strategy utilizing the REVIDE dataset are essentially better compared to different techniques. The proposed MDF strategy further develops PSNR and SSIM measurements by 1.85 dB and 0.0153, individually, contrasted with others. Thusly, the above results confirm the productivity of the proposed MDF strategy.

EXPECTED OUTPUT

The expected is the de-hazed video/image will be analyzed and spot the victims then alerted to the rescue team.

[INPUT IMAGE]

[OUTPUT IMAGE]

CONCLUSION

This study proposed a MDF outline that furnishes rigidity for managing unsettled outlines. The point extractor grounded on multi-scale consideration is utilized to prize and circuit highlights at various circumstances to give vigorous point portrayal for tape dehazing networks. A pixel AMF and a reproduction module grounded on distended swelled difficulties are utilized to more adjust different edges and recreate inactive casings, freely. far reaching preliminaries on the REVIDE dataset were led to show the viability of the MDF and its variables. Contrasted and different styles, albeit the consequences of SOTA are accomplished by MDF, there's still a little variety disfigurement. in this manner, later on, we desire to propose a variety pay medium to break the variety deformity issue, and expect to propose a more featherlight tape dehazing casing to meet the necessities of assiduity.

FUTURE WORK

Our framework probably won't function admirably for videos with very weighty dimness, and more past information is required. Likewise, however our framework accomplishes prevalent execution and quicker speed than various dehazing styles, it actually might not meet the continuous at any point interest. In this way, we improve to light weight outline with chipping away at weighty fog and continuous dehazing speed.

REFERENCES

[1] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In CVPRW, 2020. 2

[2] Mario Bijelic, Tobias Gruber, Fahim Mannan, Florian Kraus, Werner Ritter, Klaus Dietmayer, and Felix Heide. Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather. In CVPR, 2020. 3

[3] Jiezhang Cao, Yawei Li, Kai Zhang, and Luc Van Gool. Video super-resolution transformer. arXiv preprint arXiv:2106.06847, 2021. 2

[4] Kelvin CK Chan, Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Basicvsr: The search for essential components in video super-resolution and beyond. In CVPR, 2021. 2

[5] Kelvin CK Chan, Shangchen Zhou, Xiangyu Xu, and Chen Change Loy. Basicvsr++: Improving video superresolution with enhanced propagation and alignment. In CVPR, 2022. 6, 7, 8

[6] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hua. Gated context aggregation network for image dehazing and deraining. In WACV, 2019. 2

[7] Zijun Deng, Lei Zhu, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Qing Zhang, Jing Qin, and Pheng-Ann Heng. Deep multi-model fusion for single-image dehazing. In ICCV, 2019. 2, 6

[8] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In CVPR, 2020. 2, 6, 7, 8

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In ICLR, 2020. 2

[10] Vitor Guizilini, Rares Ambrus, Sudeep Pillai, Allan Raventos, and Adrien Gaidon. 3d packing for self-supervised monocular depth estimation. In CVPR, 2020. 3, 7

[11] Chun-Le Guo, Qixin Yan, Saeed Anwar, Runmin Cong, Wenqi Ren, and Chongyi Li. Image dehazing transformer with transmission-aware 3d position embedding. In CVPR, 2022. 2, 6, 7, 8

[12] Cong Huang, Jiahao Li, Bin Li, Dong Liu, and Yan Lu. Neural compression-based feature learning for video restoration. In CVPR, 2022. 2, 6

[13] Kuan-Chih Huang, Tsung-Han Wu, Hung-Ting Su, and Winston H Hsu. Monodtr: Monocular 3d object detection with depth-aware transformer. In CVPR, 2022. 4

[14] Dahun Kim, Sanghyun Woo, Joon-Young Lee, and In So Kweon. Video panoptic segmentation. In CVPR, 2020. 3, 7

[15] Johannes Kopf, Xuejian Rong, and Jia-Bin Huang. Robust consistent video depth estimation. In CVPR, 2021. 3

[16] Jingyun Liang, Yuchen Fan, Xiaoyu Xiang, Rakesh Ranjan, Eddy Ilg, Simon Green, Jiezhang Cao, Kai Zhang, Radu Timofte, and Luc Van Gool. Recurrent video restoration transformer with guided deformable attention. In NeurlPS, 2022. 2

[17] Jing Lin, Yuanhao Cai, Xiaowan Hu, Haoqian Wang, Youliang Yan, Xueyi Zou, Henghui Ding, Yulun Zhang, Radu Timofte, and Luc Van Gool. Flow-guided sparse transformer for video deblurring. In ICML, 2022. 2

[18] Ye Liu, Liang Wan, Huazhu Fu, Jing Qin, and Lei Zhu. Phase-based memory network for video dehazing. In ACM MM, 2022. 2

[19] Ye Liu, Lei Zhu, Shunda Pei, Huazhu Fu, Jing Qin, Qing Zhang, Liang Wan, and Wei Feng. From synthetic to real: