Privacy-Aware Visual Memorability in Traffic Surveillance

Yuxin Gao
Smith College
Massachusetts, USA
ygao@smith.edu

Arani Bhattacharya
IIIT-Delhi
New Delhi, India
arani@iiitd.ac.in

SouYoung Jin
Dartmouth College
New Hampshire, USA
souyoung.jin@dartmouth.edu

Shinyoung Cho
Smith College
Massachusetts, USA
scho@smith.edu

Abstract

Traffic surveillance enhances road safety and urban planning, but raises concerns that captured footage can be used to violate the privacy of drivers, passengers, or pedestrians. While current work focuses on protecting privacy from machine identifiers, the ability of human observers to memorize and recognize vehicles remains unstudied. This paper explores which vehicle features contribute to human memory retention through a pipeline designed to estimate vehicle memorability based on attention modeling and object-level features. We adopt AMNet, an attention-based memorability prediction model, to analyze how simple visual attributes—size, position, and color—affect vehicle memorability. We quantify these aspects and identify that memorability can be altered by simple changes in attributes such as vehicle's angle of view and distance from the camera. These findings offer insights for designing privacy-aware surveillance systems that address human perception threats.

1 Introduction

Traffic cameras are widely used to support traffic monitoring, road safety, and long-term urban planning [9, 21]. Despite their proven effectiveness, the ability to track individuals or vehicles in multiple camera views raises privacy concerns that limit a wider deployment [27, 28]. Advances in Multi-Target Multi-Camera Tracking (MTMCT) have intensified privacy concerns, as these techniques have become increasingly accurate at tracking targets by leveraging features such as vehicle shape across multiple camera views from different locations [10, 17, 19, 24]. In response, privacy-preserving video analytics have been extensively studied [4, 22, 29, 32, 35], focusing on obscuring privacy-sensitive regions, such as faces and license plates, or on adding noise to ensure differential privacy.

While privacy protection against MTMCT remains a critical concern, human observers present an additional yet underexplored privacy threat that operates on a different scale but warrants systematic investigation. Multi-camera traffic surveillance system involves various human stakeholders–system operators, maintenance personnel, and incidental observers–who may inadvertently compromise individual privacy (*i.e.* movement tracking or behavioral profiling), even when explicit identifiers such as faces or license plates are blurred [16]. Observers may still recognize or recall individuals based on visual characteristics such as unique colors and bumper stickers [1]. For instance, a vintage red convertible with bumper stickers in a small town remains easily recognizable, and may become unintentionally memorable, creating persistent privacy risks. This phenomenon aligns with prior work in cognitive science, which shows that human memory is strongly influenced by salience and visual attention [31].

We investigate whether certain vehicles exhibit greater memorability than others in traffic surveillance footage and seek to identify the object-level visual factors, such as size, position, and color, that contribute to differential memorability. To the best of our knowledge, visual memorability as a privacy-relevant attribute remains underexplored in the surveillance literature. We present a pipeline to estimate object-level memorability from surveillance footage and attention modeling[1]. We found that such simple visual features significantly influence whether a vehicle is likely to be remembered by human observers. Furthermore, even changing the camera's angle and distance from the streets can significantly alter the memorability, with vehicles having a frontal view, appearing on the lower part of the view and closer to the camera having significantly higher memorability than other vehicles.

2 Our Threat Model

We consider honest-but-curious human observers as our threat model. These observers, such as system operators, have legitimate access to traffic surveillance videos. We assume that observers rely on visual perception and memory rather than automated systems (*e.g.,* MTMCT) to recognize vehicles across time and space.

This threat model is grounded in established cognitive research. Studies have shown that visually salient stimuli—those that stand out due to contrast, distinctiveness, or semantic content—are more likely to capture attention and be remembered [31]. Attention plays a key role in whether incidental stimuli are encoded into long-term memory, with people able to recall thousands of images after a single glance [2]. Memorability is a measurable and consistent image property. Certain object-level cues (*e.g.,* size, color, and positioning) significantly influence retention across viewers [13].

[1]To facilitate reproducibility, code repository for this paper is accessible at: https://github.com/EvelynGao233/Privacy-Aware-Visual-Memorability-in-Traffic-Surveillance

WPES, Taipei, Taiwan

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

These perceptual mechanisms create privacy risks in traffic surveillance. Vehicles with uncommon colors, large footprints, or unique modifications may become memorable and vulnerable to informal re-identification. Over time, observers may associate such vehicles with specific routines, locations, or individuals. While existing privacy-preserving approaches aim to prevent machine-based recognition through blurring, encryption, or differential privacy, they often overlook vulnerabilities from human perception. Our threat model addresses this gap by focusing on re-identification that occurs even without malicious intent or technical aid.

3 Background

Visual memorability refers to how likely a visual stimulus is to be retained after a brief encounter. Foundational studies have shown that memorability is not entirely subjective but remains consistent across viewers and can be reliably predicted by computational models [3, 8, 13]. This has led to a growing interest in treating memorability as a quantifiable and intrinsic property of visual content. AMNet [8] demonstrated that incorporating an attention mechanism into a deep neural network significantly improves memorability prediction performance. Trained on a dataset of over 60,000 natural images with behavioral recall-based memorability labels, AMNet enables quantification of which objects are most likely to be remembered by human observers. We extend this foundation to explore whether mosaic stylization (e.g., neural style transfer [20]) can reduce vehicle memorability in surveillance contexts. This approach is central to our study of perception-driven privacy risks in traffic surveillance.

4 Data Processing Pipeline

We implement an attention-based memorability estimation pipeline to quantify object-level visual memorability in urban traffic surveillance footage (Figure 1).

Step 1: Multi-Camera Vehicle Tracking and Identification. Our pipeline begins by identifying vehicles across multiple cameras using a state-of-the-art framework [17]. This framework integrates vehicle detection, intra-camera tracking, and cross-camera association to produce standardized annotations. Each vehicle detection includes an object ID, source camera ID, frame ID, and bounding box coordinates. These annotations provide the temporal and spatial foundation for subsequent memorability analysis.

Step 2: Per-Frame Object Memorability Estimation. We apply AMNet [8] to generate spatial attention maps for each frame, indicating which regions most strongly contribute to the predicted memorability score. To derive object-level memorability, we align each vehicle's bounding box with the corresponding attention map, scale coordinates to match the attention map's resolution, extract the overlapping subregion, and compute the mean attention intensity within each vehicle region. This value serves as a proxy for the object's memorability. Repeating this procedure across all vehicle instances and frames yields memorability scores indexed by object ID and time.

Step 3: Object-Level Visual Feature Extraction. To examine which visual properties are associated with higher memorability, we extract three interpretable object-level features for each vehicle instance: *size*, *position*, and *dominant color*.

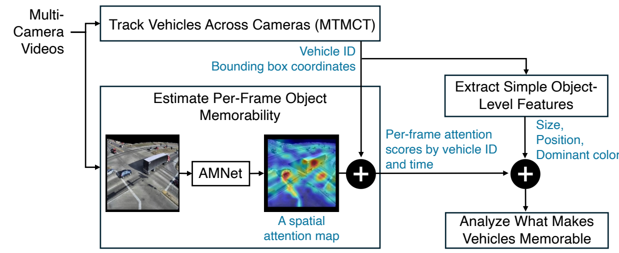

Figure 1: Object-level memorability assessment pipeline

Size is calculated as the bounding box area ($Area = w \times h$), which captures the relative prominence of the object in the frame. *Position* is encoded using normalized center coordinates (cx, cy). The vertical coordinate cy is particularly relevant in fixed-camera surveillance, as objects near the bottom of the frame are typically closer to the camera and more visually salient. *Dominant color* is determined using an off-the-shelf color classifier [26] that assigns categorical color labels to vehicle patches.

Step 4: Constructing the Final Dataset. We integrate the MTMC tracking outputs, per-frame object memorability scores, and visual features to create a unified object-level dataset. Each record contains tracking information, spatial features (area and position), color, and memorability score for each vehicle-frame pair. This enables statistical analysis to identify which visual attributes contribute most significantly to perceived memorability in surveillance footage.

5 Analysis

We apply our analysis on a multi-video multi-tracking dataset from the NVIDIA AI City Challenge [23], a benchmark for research in intelligent video analysis. The dataset comprises traffic camera videos from a U.S. city with license plates blurred to protect identity. These real-world surveillance videos are well-suited for studying how visual properties influence memorability in practical settings.

5.1 Geometric Feature Analysis

We examine correlations between geometric features and memorability scores. Figure 2 shows Spearman correlations for center coordinates (cx, cy), $area$, and object count (discussed in Section 5.3). The results indicate that object area has the strongest correlation with memorability ($area$, $\rho = 0.55$), followed by vertical position (cy, $\rho = 0.39$), consistent with research showing that size and foreground prominence enhance attention and recall [33]. In contrast, horizontal position (cx, $\rho = -0.02$) shows weak correlation. This analysis confirms that larger and lower-positioned vehicles are more memorable, likely due to their visual salience and proximity to the camera.

5.2 Color Distinctiveness Analysis

We examine how memorability varies across vehicle colors by computing average AMNet attention scores and occurrence counts for each color label. The scatterplot (Figure 3) reveals an inverse pattern: less common vehicle colors (yellow, brown, silver) exhibit higher attention scores, while common colors (black, white) show lower scores. This pattern suggests that visual distinctiveness contributes to salience, with uncommon colors attracting more attention. A

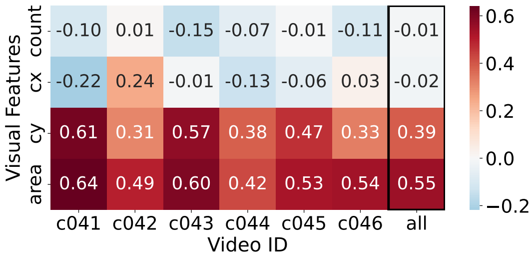

Figure 2: Correlation between visual features and memorability scores across videos. The 'all' column (black border) represents overall correlations across the combined dataset

one-way ANOVA (Analysis of Variance) confirms significant differences across color categories ($F = 70.55$, $p \ll 0.001$), supporting color's inclusion as a predictor in subsequent analyses.

Limitation. These patterns should be interpreted with caution, as color labels from a pre-trained classifier [26] may not capture perceptual features such as brightness, glossiness, or saturation (*e.g.*, orange objects appearing darker).

Figure 3: Mean AMNet attention score (memorability proxy), plotted against detection frequency for each vehicle color

5.3 Multivariate Memorability Modeling

To quantify relationships between visual features and predicted memorability, we develop five hierarchical linear regression models progressively adding feature groups: object size (*area*), spatial position (*cx*, *cy*), visual appearance (*color*), and object density (*objectCount*), with interaction terms to capture compound effects. Table 1 shows that each step yields incremental improvements in adjusted R^2. The substantial improvement from Model 2 to Model 3 ($\Delta R^2 = 0.052$) demonstrates synergistic size-position interactions beyond independent effects. Models 4 and 5 achieve the best performance (adjusted $R^2 = 0.279$), which is reasonable given the memorability's inherent variability—factors like motion, prior exposure, or scene semantics remain uncaptured. The identical performance between these models indicates that object count (the number of detected objects per frame) provides no additional predictive power beyond geometric and color features, consistent with weak correlations observed in Figure 2 ($\rho = -0.01$ overall).

5.4 Viewpoint Effects on Memorability

While object-level features predict visual memorability, camera placement and scene layout may also influence attention. We compare six videos (c041–c046) from distinct surveillance viewpoints at a multi-lane intersection (Figure 4).

Table 1: Comparison of linear regression models for predicting memorability scores. Higher adjusted R^2 indicates better fit. All models significant at $p < 0.001$.

Model	Formula	Adj R^2
Model 1	$area$	0.184
Model 2	$area + cx + cy$	0.202
Model 3	$area * cx * cy$	0.254
Model 4	$area * cx * cy + \mathrm{C}(color)$	**0.279**
Model 5	$area * cx * cy + \mathrm{C}(color) + objectCount$	**0.279**

Figure 4: Representative frames from six surveillance videos

Videos from frontal or lower camera angles (c041, c043) yield higher average memorability scores, while top-down perspectives (c044) produce lower memorability (Figure 5). AMNet attention maps confirm that vehicles in c044 receive little predicted focus despite remaining visually present, with attention broadly dispersed across background regions, unlike c046 where attention focuses on vehicles (Figure 6). Analysis shows vehicles in c044 consistently appear higher in the frame (*cy*), coinciding with reduced attention scores (Figure 7). Since human attention favors center or lower image regions, this positioning may reduce vehicle salience and memorability.

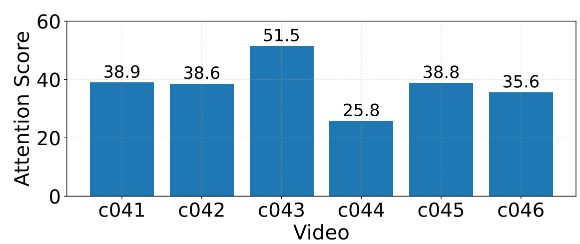

Figure 5: Mean memorability scores across the six videos. c043 has the highest average, while c044 is substantially lower

Cross-view analysis reveals substantial memorability variation for the same vehicles across different viewpoints. 75.4% of multi-camera objects show standard deviations greater than 10 points, and 49.1% vary by more than 30 points across views. This substantial variation demonstrates that memorability depends on the interaction between object attributes and viewing context. Camera angle influences the apparent object size and spatial positioning we identified as key factors, while background context and framing create additional scene-level effects that could be leveraged for privacy enhancement.

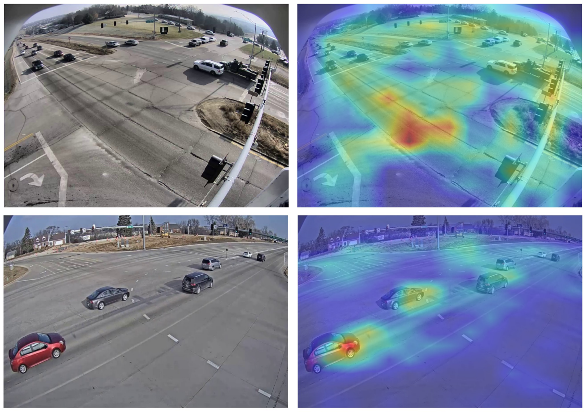

Figure 6: Comparison of AMNet attention patterns between videos c044 (top) and c046 (bottom). The heatmaps visualize predicted visual salience with red indicating high attention and blue indicating low attention. Video c044 shows dispersed attention across background elements, while c046 exhibits focused attention on vehicles

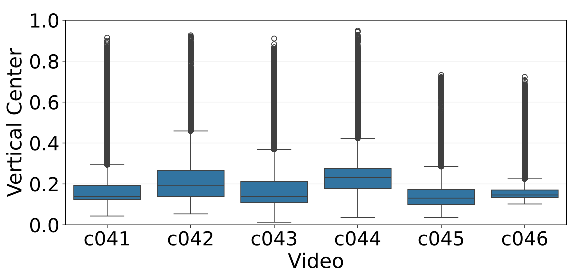

Figure 7: Normalized vertical center (cy) of objects across videos. Vehicles in c044 appear higher in the frame on average, which may affect visual salience

6 Related Work

Visual Memorability. Visual memorability refers to the likelihood of visual content being retained in human memory after brief exposure. Studies show that memorability is consistent across viewers and can be computationally predicted [3, 8, 13]. Recent works demonstrate that object-level features—like size, position, and visual distinctiveness—influence memorability [14, 15]. For instance, centrally positioned or larger objects tend to be more memorable than peripheral or smaller ones [11]. Memento [25], VideoMem [5] and Modular Memorability [7] similarly identify both low-level and high-level features of videos that allow us to predict their memorability. Our work specifically analyzes the memorability for traffic surveillance objects, using the predictions of AMNet.

Privacy in Traffic Surveillance. Current privacy-preserving approaches focus on defending against algorithmic threats such as license plate recognition and vehicle re-identification [4, 18]. Common mitigation strategies include face and plate de-identification [6,

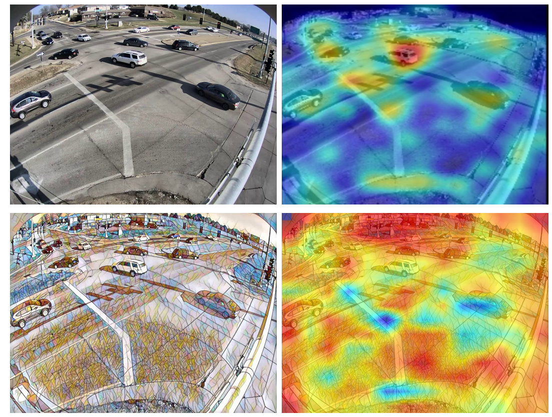

Figure 8: Effect of mosaic stylization on attention patterns. The top row shows the original frame and its attention map; the bottom row shows the same frame with mosaic stylization and its attention map. The stylized frame produces more diffuse attention, reducing focus on vehicles

12], differential privacy [4], and video anonymization [22, 32]. However, unlike our work. these techniques primarily address machine-based recognition while overlooking human perception threats.

Multi-Camera Tracking. Modern traffic surveillance relies on multi-target multi-camera (MTMC) tracking to monitor vehicles across distributed camera networks [17, 30]. Recent advances like FairMOT [34] unify detection and identity embedding for improved performance, with frameworks validated in benchmarks such as the AI City Challenge [24]. These recent developments in computer vision, while useful to improve traffic safety and optimization, also lead to increase in threats to privacy of citizens by deanonymizing their locations. Our work attempts to mitigate such problems.

7 Conclusion

In this work, we addressed human perception threats in traffic surveillance privacy through an attention-based memorability pipeline. We demonstrated that simple visual features, such as size, position, and color, significantly influence vehicle memorability, with substantial variation across camera viewpoints. Our regression analysis revealed that geometric interactions (*area * position*) are crucial for memorability prediction, while object density shows minimal impact. Preliminary experiments with mosaic stylization suggest visual transformations can redistribute attention patterns, providing a foundation for privacy-preserving surveillance approaches.

For future work, we propose investigating how image transformations can alter memorability to enhance privacy, while preserving public safety utility. Our preliminary experiments with mosaic stylization demonstrate that visual transformations can indeed redistribute attention patterns and reduce vehicle memorability (Figure 8), but we intend to perform more extensive experiments. We further plan to investigate whether full-frame visual style transfer and partial zooming techniques can reduce object-level memorability by redistributing attention across scenes. User studies and eye-tracking experiments would provide valuable validation of these attention-based proxies in real-world perceptual contexts.

Acknowledgments

We thank Dr. Yael Granot and Dr. Preston Thakral for their generous time and expertise in experimental psychology and memory research, offering valuable perspectives on human perception and cognitive processes that enriched our memorability analysis.

References

[1] Ashutosh Holla B., Manohara M. M. Pai, Ujjwal Verma, and Radhika M. Pai. 2025. Vehicle Re-Identification and Tracking: Algorithmic Approach, Challenges and Future Directions. *IEEE Open Journal of Intelligent Transportation Systems* 6 (2025), 155–183. doi:IEEE 10.1109/OJITS.2025.3538037

[2] Timothy Brady, Talia Konkle, George Alvarez, and Aude Oliva. 2008. Visual long-term memory has a massive capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America* 105 (10 2008), 14325–9. doi:10.1073/pnas.0803390105

[3] Zoya Bylinskii, Lore Goetschalckx, Anelise Newman, and Aude Oliva. 2022. *Memorability: An Image-Computable Measure of Information Utility.* Springer International Publishing, Cham, 207–239. doi:10.1007/978-3-030-81465-6_8

[4] Frank Cangialosi, Neil Agarwal, Venkat Arun, Srinivas Narayana, Anand Sarwate, and Ravi Netravali. 2022. Privid: Practical, Privacy-Preserving Video Analytics Queries. In *19th USENIX Symposium on Networked Systems Design and Implementation (NSDI 22)*. USENIX Association, Renton, WA, 209–228. https://www.usenix.org/conference/nsdi22/presentation/cangialosi

[5] Romain Cohendet, Claire-Helene Demarty, Ngoc Duong, and Martin Engilberge. 2019. VideoMem: Constructing, Analyzing, Predicting Short-Term and Long-Term Video Memorability. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 2531–2540. doi:10.1109/ICCV.2019.00262

[6] Liang Du and Haibin Ling. 2011. Preservative License Plate De-identification for Privacy Protection. In *2011 International Conference on Document Analysis and Recognition*. 468–472. doi:10.1109/ICDAR.2011.100

[7] Théo Dumont, Juan Segundo Hevia, and Camilo L. Fosco. 2023. Modular Memorability: Tiered Representations for Video Memorability Prediction. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10751–10760. doi:10.1109/CVPR52729.2023.01035

[8] Jiri Fajtl, Vasileios Argyriou, Dorothy Monekosso, and Paolo Remagnino. 2018. AMNet: Memorability Estimation with Attention. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6363–6372. doi:10.1109/CVPR.2018.00666

[9] Aleksandr Fedorov, Kseniia Nikolskaia, Sergey Ivanov, Vladimir Shepelev, and Alexey Minbaleev. 2019. Traffic flow estimation with data from a video surveillance camera. *Journal of Big Data* 6, 1 (2019), 73. doi:10.1186/s40537-019-0234-z

[10] Lunlin Fei and Bing Han. 2023. Multi-Object Multi-Camera Tracking Based on Deep Learning for Intelligent Transportation: A Review. *Sensors* 23, 8 (2023), 3852. doi:10.3390/s23083852

[11] Lore Goetschalckx, Alex Andonian, Aude Oliva, and Phillip Isola. 2019. GANalyze: Toward Visual Definitions of Cognitive Image Properties. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 5743–5752. doi:10.1109/ICCV.2019.00584

[12] R. Gross, L. Sweeney, F. de la Torre, and S. Baker. 2006. Model-Based Face De-Identification. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)*. 161–161. doi:10.1109/CVPRW.2006.125

[13] Phillip Isola, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2011. What makes an image memorable?. In *IEEE Computer Vision and Pattern Recognition*. 145–152. doi:10.1109/CVPR.2011.5995721

[14] Aditya Khosla, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2012. Memorability of image regions. In *Advances in neural information processing systems*. 296–304.

[15] Joonyeon Kim, Seunghoon Yoon, and Vladimir Pavlovic. 2013. Relative spatial features for image memorability. In *Proceedings of the 21st ACM international conference on Multimedia*. 761–764. doi:10.1145/2502081.250219

[16] Lamyanba Laishram, Muhammad Shaheryar, Jong Taek Lee, and Soon Ki Jung. 2025. Toward a Privacy-Preserving Face Recognition System: A Survey of Leakages and Solutions. *ACM Comput. Surv.* 57, 6, Article 147 (Feb. 2025), 38 pages. doi:10.1145/3673224

[17] Chong Liu, Yuqi Zhang, Hao Luo, Jiasheng Tang, Weihua Chen, Xianzhe Xu, Fan Wang, Hao Li, and Yi-Dong Shen. 2021. City-Scale Multi-Camera Vehicle Tracking Guided by Crossroad Zones. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 4124–4132. doi:10.1109/CVPRW53098.2021.00466

[18] Hao Luo, Weihua Chen, Xianzhe Xu, Jianyang Gu, Yuqi Zhang, Chong Liu, Yiqi Jiang, Shuting He, Fan Wang, and Hao Li. 2021. An Empirical Study of Vehicle Re-Identification on the AI City Challenge. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 4090–4097. doi:10.1109/CVPRW53098.2021.00462

[19] Wenhan Luo, Junliang Xing, Anton Milan, Xiaoqin Zhang, Wei Liu, and Tae-Kyun Kim. 2021. Multiple object tracking: A literature review. *Artificial Intelligence* 293 (2021), 103448. doi:10.1016/j.artint.2020.103448

[20] Rusty Mina. 2018. fast-neural-style: Fast Style Transfer in Pytorch! https://github.com/iamRusty/fast-neural-style-pytorch.

[21] Yanjinlkham Myagmar-Ochir and Wooseong Kim. 2023. A Survey of Video Surveillance Systems in Smart City. *Electronics* 12, 17 (2023). doi:10.3390/electronics12173567

[22] Sowmya Myneni, Garima Agrawal, Yuli Deng, Ankur Chowdhary, Neha Vadnere, and Dijiang Huang. 2022. SCVS: On AI and Edge Clouds Enabled Privacy-preserved Smart-city Video Surveillance Services. *ACM Transactions on Internet of Things* 3, 4 (2022), 28:1–28:26. doi:10.1145/3542953

[23] Milind Naphade, David C. Anastasiu, Anuj Sharma, Vamsi Jagrlamudi, Hyeran Jeon, Kaikai Liu, Ming-Ching Chang, Siwei Lyu, and Zeyu Gao. 2017. The NVIDIA AI City Challenge. In *2017 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. 1–6. doi:10.1109/UIC-ATC.2017.8397673

[24] Milind Naphade, Shuo Wang, David C. Anastasiu, Zheng Tang, Ming-Ching Chang, Xiaodong Yang, Yue Yao, Liang Zheng, Pranamesh Chakraborty, Anuj Sharma, Qi Feng, Vitaly Ablavsky, and Stan Sclaroff. 2021. The 5th AI City Challenge. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 4258–4268. doi:10.1109/CVPRW53098.2021.00482

[25] Anelise Newman, Camilo Fosco, Vincent Casser, Allen Lee, Barry McNamara, and Aude Oliva. 2020. Multimodal Memorability: Modeling Effects of Semantics and Decay on Video Memorability. In *Computer Vision – ECCV 2020*. Springer International Publishing, Cham, 223–240. doi:10.1007/978-3-030-58517-4_14

[26] Nikalosa. July 23, 2020. Vehicle Make and Color Recognition. https://github.com/nikalosa/Vehicle-Make-Color-Recognition. Accessed: 2025-04-17.

[27] Jacob Posik. Apri. Big Brother Wants to Watch You on Your Daily Commute. https://www.themainewire.com/2025/04/big-brother-wants-to-watch-you-on-your-daily-commute/.

[28] Tony Saavedra. June. Police surveillance using private security cameras in real time stirs 'Big Brother' fears. https://www.dailynews.com/2023/06/16/police-surveillance-using-private-security-cameras-in-real-time-stirs-big-brother-fears/. Accessed on July 17, 2023.

[29] Mehwish Tahir, Yuansong Qiao, Nadia Kanwal, Brian Lee, and Mamoona Naveed Asghar. 2023. Privacy Preserved Video Summarization of Road Traffic Events for IoT Smart Cities. *Cryptography* 7, 1 (2023). doi:10.3390/cryptography7010007

[30] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birchfield, Shuo Wang, Ratnesh Kumar, David Anastasiu, and Jenq-Neng Hwang. 2019. CityFlow: A City-Scale Benchmark for Multi-Target Multi-Camera Vehicle Tracking and Re-Identification. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 8789–8798. doi:10.1109/CVPR.2019.00900

[31] Benjamin Tatler, Mary Hayhoe, M. Land, and Dana Ballard. 2011. Eye guidance in natural vision: Reinterpreting salience. *Journal of vision* 11 (05 2011), 5. doi:10.1167/11.5.5

[32] Han Wang, Shangyu Xie, and Yuan Hong. 2020. Videodp: A flexible platform for video analytics with differential privacy. *Proceedings on Privacy Enhancing Technologies* (2020), 277–296. Issue 4. doi:10.2478/popets-2020-0073

[33] Jeremy M. Wolfe and Todd S. Horowitz. 2021. Five Factors That Guide Attention in Visual Search. *Nature Human Behaviour* 5 (2021), 289–299. doi:10.1038/s41562-017-0058

[34] Yifu Zhang, Chunyu Wang, Xinggang Wang, Wenjun Zeng, and Wenyu Liu. 2021. FairMOT: On the Fairness of Detection and Re-identification in Multiple Object Tracking. *International Journal of Computer Vision* 129 (2021), 3069–3087. doi:10.1007/s11263-021-01513-4

[35] Ruoyu Zhao, Yushu Zhang, Tao Wang, Wenying Wen, Yong Xiang, and Xiaochun Cao. 2024. Visual Content Privacy Protection: A Survey. *Comput. Surveys* 57, 5 (2024), 1–35. doi:10.1145/3708501