Physics-Based modeling for Synthetic Urban Flood Rendering from Monocular Video

MD Fayaz Bin Hossen, Ahmed Temtam, Kwame AMPOFO, and Khan M. Iftekharuddin
Department of Electrical and Computer Engineering, Vision Lab
Old Dominion University, Norfolk, Virginia, USA
{mhoss006, atemt001, kampo001, kiftekha}@odu.edu

***Abstract*—Reliable reconstruction of urban flooding is important for urban planning, flood monitoring, and risk analysis; however, progress is hindered by the limited availability of datasets with usable scene geometry and flood-depth annotations. Existing approaches often rely on LiDAR scanning, 3D reconstruction, or purely generative models, which can be computationally expensive and may not enforce physically realistic fluid dynamics over time. Moreover, obtaining labeled ground truth data —particularly flood depth —is costly and difficult to scale. To address these limitations, we propose a monocular-video-driven framework that combines open-vocabulary semantic segmentation, monocular depth estimation, and a lightweight physics-based simulation model using Shallow Water Equations (SWE). The simulation operates on scene-derived terrain and dynamic obstacles to produce temporally consistent flooding evolution, which is then used to synthesize flooding imagery and corresponding dense supervision signals. As a proof of concept, we also demonstrate a physics-informed neural surrogate model that regularizes predictions with SWE residuals. Overall, the proposed framework enables efficient generation of physically grounded synthetic data for training and benchmarking flood segmentation and depth estimation models, supporting data-driven urban resilience and disaster preparedness.**

***Index Terms*—Urban Flood Modeling, Shallow Water Equations (SWE), Semantic Segmentation, Depth Estimation, Synthetic Data Generation**

I. INTRODUCTION

Flooding remains one of the most devastating natural disasters globally, driven by interactions between hydrological hazards—such as riverine flooding from heavy precipitation—and societal vulnerabilities, including infrastructure exposure and population density [1]. Research indicates that rising sea levels will worsen this threat, with projections suggesting that by 2050 many coastal communities in the United States will experience flooding on thirty or more days annually. As weather-related catastrophes become more frequent, effective flood risk management has become a global priority.

To assess danger and mitigate damage, image data from satellites, unmanned aerial vehicles (drones), and surveillance cameras has become indispensable. While early work focused on remote sensing via satellite imagery [2], these methods often lack the street-level detail needed for immediate urban response, making camera-based computer vision a practical complement to traditional sensor deployments [3].

Despite this potential, computer vision for flood monitoring is limited by the lack of high-quality, actionable data. Widely used flood datasets are largely image-centric and lack the temporal continuity needed for traffic-camera monitoring and depth-aware learning [4], [5]. Recent works using traffic imagery and fixed cameras often require scene-specific calibration and still face limited large-scale depth supervision under challenging conditions such as variable lighting, reflections, and water ripples [6].

To address data scarcity, generative models have been proposed for synthetic flood imagery and video. However, purely generative approaches often struggle to maintain physical realism and do not provide reliable pixel-wise flood depth labels [7], while LiDAR-based depth and 3D reconstruction remain expensive for large-scale dataset generation.

To overcome these limitations, we propose a framework that integrates semantic segmentation, depth estimation, and physics-based simulation using the Shallow Water Equations (SWE). Unlike methods based solely on style transfer or inpainting, our approach uses physically grounded dynamics to simulate realistic flooding on urban terrain, enabling efficient generation of annotated data with flood masks and depth maps. Section II reviews relevant background, Section III presents the pipeline, Section IV reports experimental results, and Section V concludes with limitations and future directions.

II. BACKGROUND

A. Related Work

The detection of floodwater and estimation of water levels have primarily relied on deep learning models applied to real-world imagery [8]–[11]. Research in this area has used feature extraction and pre-processing to handle diverse data sources, ranging from social media images [9] to crowdsourced roadway imagery [10]. While several studies estimate flood-related water levels from images using scene cues and reference information [9], [11], these approaches remain constrained by the scarcity of high-quality labeled data, particularly paired flooded and non-flooded observations of the same location.

To mitigate limited labels, researchers have increasingly explored synthetic data generation. GAN-based methods have been used to produce synthetic flood imagery, including ClimateGAN-style approaches for static transformations [12]–[14]. For video monitoring scenarios, Lamczyk et al. [7] demonstrated synthetic flood video generation using region-based conditioning and learned appearance synthesis. However, purely generative approaches typically do not enforce

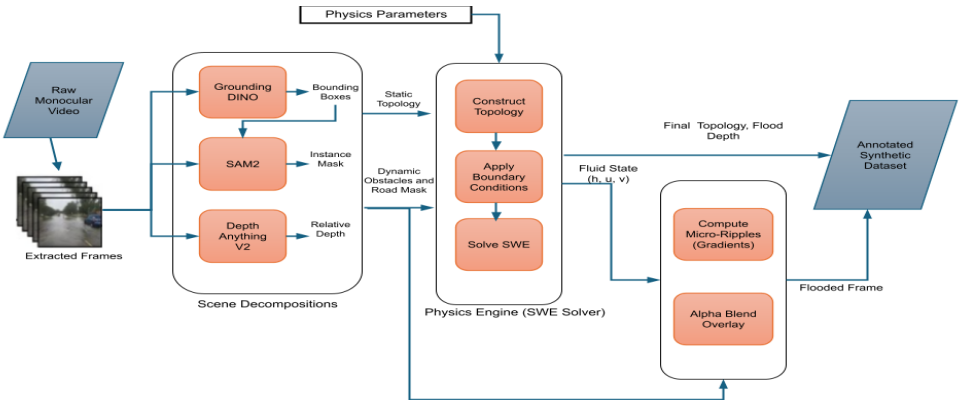

Fig. 1. Overview of the Proposed Neuro-Physical Framework.

physical constraints, making it difficult to guarantee consistency in depth, flow, and propagation behavior.

Recent efforts have aimed to improve physical fidelity through explicit simulation. Rahman et al. [15] introduced a 3D digital-twin pipeline integrating scene reconstruction and high-fidelity fluid simulation to obtain physically consistent dynamics and depth, but such methods are computationally expensive and difficult to scale. Our method addresses this gap by combining semantic segmentation with SWE-based simulation to generate scalable, physically grounded, and annotated flood datasets.

B. Semantic Segmentation and Object Detection

To accurately simulate flooding in an urban environment, it is essential to distinguish static terrain regions (e.g., roads), where water accumulates, from dynamic obstacles (e.g., vehicles) that interact with the flow. For this task, we employ Grounding DINO [16] and the Segment Anything Model 2 (SAM 2) [17]. Grounding DINO is an open-set, text-conditioned object detector that supports language-grounded detection (e.g., prompting "road" and "vehicle"). We use it to localize road and vehicle regions, producing bounding boxes that serve as prompts for SAM 2.

SAM 2 is a promptable segmentation foundation model for image and video settings; in video, it maintains object information over time for temporally coherent mask propagation. In our pipeline, we use SAM 2 with detection prompts and simple temporal stabilization, such as aggregating road masks over sampled frames and enforcing connectedness constraints, to reduce flicker and avoid non-physical discontinuities in the downstream simulation.

C. Monocular Depth Estimation

Simulating realistic water transport requires more than 2D segmentation; it also needs a scene geometry proxy to define gravity-driven flow directions. Because our setting is monocular video, we use Depth Anything [18], a foundation model for monocular depth prediction, and use its relative depth maps as a scene-consistent terrain proxy after normalization and road-region conditioning rather than as metrically calibrated bathymetry.

In our pipeline, this depth-derived terrain proxy provides the large-scale road slope used by the shallow-water solver to determine preferential flow direction, while segmentation defines where water can accumulate and propagate.

D. Physics Simulation

To generate physically grounded flood dynamics rather than purely visual approximations, we use a depth-averaged flow model based on the 2D shallow water equations (SWE) under the shallow-flow assumption. The 2D SWE are:

$$\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = 0, \tag{1}$$

$$\frac{\partial (hu)}{\partial t} + \frac{\partial}{\partial x}\left(hu^2 + \frac{1}{2}gh^2\right) + \frac{\partial (huv)}{\partial y} = -gh\frac{\partial B}{\partial x}, \tag{2}$$

$$\frac{\partial (hv)}{\partial t} + \frac{\partial}{\partial y}\left(hv^2 + \frac{1}{2}gh^2\right) + \frac{\partial (huv)}{\partial x} = -gh\frac{\partial B}{\partial y}. \tag{3}$$

Here, h is water depth, u and v are velocities, B is bottom topography, and g is gravitational acceleration.

In our work, vehicle masks are treated as impermeable obstacles that prevent water accumulation inside vehicle regions and damp flow near their boundaries, producing realistic local wake and deflection effects absent in purely generative flood overlays.

E. Physics-Informed Neural Surrogate Model

To complement explicit SWE time-stepping, we employ a physics-informed neural surrogate as a proof of concept. The surrogate takes the constructed terrain proxy as input and predicts spatiotemporal fields (u, v, η) over a fixed horizon, trained using data supervision (when available) and SWE residual regularization. This follows the idea of physics-informed learning, popularized by Physics-Informed Neural Networks (PINNs), where physical laws constrain the hypothesis space beyond purely data-driven fitting.

Instead of solving the SWE iteratively at inference time, the surrogate amortizes this computation into a single forward pass, enabling fast rollout for video-scale generation. This is consistent with recent physics-informed operator learning approaches that efficiently learn PDE solution maps while preserving physical structure.

For training and sanity checks, we use publicly available hydrodynamic forecast outputs (e.g., Great Lakes forecasting products) as one source of physically meaningful velocity and free-surface targets, then condition downstream rendering on the surrogate-predicted motion fields.s, and then condition downstream rendering on the surrogate-predicted motion fields.

III. METHOD

We propose a neuro-physical framework for generating synthetic flood datasets by explicitly modeling the interaction between urban geometry and fluid dynamics. The pipeline defines

$$F : \mathcal{V} \to D_{\text{syn}}, \tag{4}$$

where a monocular video sequence $\mathcal{V}$ is transformed into a synthetic dataset D_{syn} containing synchronized video, segmentation masks, and pixel-wise depth maps. The architecture comprises four coupled modules: (A) Spatiotemporal Scene Decomposition, (B) Topographic Surface Reconstruction, (C) Differentiable Fluid Simulation, and (D) Photorealistic Rendering & Annotation (Fig. 1).

A. Spatiotemporal Scene Decomposition

Given a raw monocular street-level video, we decompose each sequence into static terrain regions (roadways), dynamic obstacles (vehicles), and a geometry proxy. We initialize scene understanding using Grounding DINO [16]; given a frame I_t, zero-shot text constraints $T = \{\text{“car”, “truck”, “bus”}\}$ retrieve bounding box proposals B_t without retraining.

We then use SAM 2 [17] to produce prompt-guided segmentation masks. Because frame-wise flicker can destabilize downstream fluid evolution, we apply simple temporal stabilization, including aggregation across sampled frames, largest-component filtering, and morphological closing, to obtain

$$M_{\text{road}} \in \{0,1\}^{H\times W}\text{: road mask,} \tag{5}$$

$$M_{\text{veh}} \in \{0,1\}^{H\times W}\text{: vehicle obstacle mask.} \tag{6}$$

These masks define the physical domain and obstacle constraints for the physics module.

B. Topographic Surface Reconstruction

The Shallow Water Equations require a bottom topography field $B(x, y)$ to determine gravity-driven flow. Since LiDAR is often unavailable for crowd-sourced videos, we approximate $B(x, y)$ using monocular depth estimation. We employ Depth Anything V2 [18] to extract a relative depth map $D_{\text{rel}}(x, y)$, treated as a scene-consistent terrain proxy after normalization and road conditioning. To encourage realistic urban drainage, we add a weak road-crown prior:

$$B(x, y) = \psi(D_{\text{rel}}(x, y)) + \alpha\,\text{EDT}(M_{\text{road}}), \tag{7}$$

where $\psi(\cdot)$ maps relative depth to normalized elevation, $\text{EDT}(\cdot)$ is the Euclidean distance transform on the road mask, and α controls crown amplitude. This improves drainage directionality without full 3D reconstruction.

C. Physics Integration

1) Primary engine: SWE-based shallow-water solver: The primary physical engine solves a depth-averaged flow model based on the 2D Shallow Water Equations (SWE). We simulate water depth $h(x, y, t)$ and depth-averaged velocities $(u(x, y, t), v(x, y, t))$ over a regular grid.

1) Governing Equations: We solve the conservative SWE for fluid height $h(x, y, t)$ and depth-averaged discharge fluxes $\phi = (hu, hv)$:

$$\frac{\partial h}{\partial t} + \frac{\partial (hu)}{\partial x} + \frac{\partial (hv)}{\partial y} = S_{\text{rain}}, \tag{8}$$

$$\frac{\partial (hu)}{\partial t} + \frac{\partial}{\partial x}\left(hu^2 + \frac{1}{2}gh^2\right) + \frac{\partial (huv)}{\partial y} = -gh\frac{\partial B}{\partial x} - \gamma(hu), \tag{9}$$

$$\frac{\partial (hv)}{\partial t} + \frac{\partial}{\partial y}\left(hv^2 + \frac{1}{2}gh^2\right) + \frac{\partial (huv)}{\partial x} = -gh\frac{\partial B}{\partial y} - \gamma(hv), \tag{10}$$

where g is gravitational acceleration, γ is a linear damping coefficient for surface friction, and $\frac{\partial B}{\partial x}$, $\frac{\partial B}{\partial y}$ are the topographic gradients driving flow.

2) Numerical Solution: We discretize the domain Ω on a regular Cartesian grid with spatial step Δx and time step Δt, and use a finite difference method on a staggered Arakawa C-grid to reduce checkerboard pressure artifacts. At each time step, we first compute topographic slope and surface gradients using central differences, e.g.,

$$\frac{\partial B}{\partial x} \approx \frac{B_{i+1,j} - B_{i-1,j}}{2\Delta x}, \tag{11}$$

with total surface elevation $\eta = h + B$ defining the hydrostatic pressure gradient. We then update the discharge fluxes:

$$(hu)^{t+1}_{i,j} = (hu)^t_{i,j} - \Delta t\left[g\,h^t_{i,j}\left(\frac{\eta^t_{i+1,j} - \eta^t_{i-1,j}}{2\Delta x}\right) + \gamma(hu)^t_{i,j}\right], \tag{12}$$

and enforce mass conservation to update water height:

$$h^{t+1}_{i,j} = h^t_{i,j} + S_{\text{rain}} - \Delta t\left[\frac{(hu)^{t+1}_{i+1,j} - (hu)^{t+1}_{i,j}}{\Delta x} + \frac{(hv)^{t+1}_{i+1,j} - (hv)^{t+1}_{i,j}}{\Delta y}\right]. \tag{13}$$

We additionally impose

$$h^{t+1} = \max(h^{t+1}, 0) \tag{14}$$

to ensure non-negative water depth.

3) Boundary conditions: The masks from Module A are injected as constraints: water is permitted primarily on the road domain M_{road}; vehicles $M_{\text{veh}}(t)$ act as impermeable obstacles, suppressing water depth inside vehicle pixels and damping velocities near obstacle boundaries; and raised terrain

near image borders can be used as a "bowl" to reduce unphysical mass loss. This yields physically plausible deflection and wake behavior without claiming exact wall-boundary enforcement.

2) Physics-informed neural surrogate:

a) Training Data Generation and Preprocessing: We prepare training pairs (X, Y) from hydrodynamic NetCDF forecasts [19], where $X \in \mathbb{R}^{1\times H\times W}$ is static bathymetry/depth h (meters, positive down) and $Y \in \mathbb{R}^{T\times 3\times H\times W}$ contains (u, v, ζ) over time. We extract h and ζ, and use depth-averaged velocities (u_a, v_a) when available, otherwise surface-layer (u, v) at the first vertical level. Model node coordinates $(\mathrm{lon}, \mathrm{lat})$ are projected to a local planar system in meters, invalid nodes ($h \le 0$ or non-finite) are removed, and a bounded subset is chosen by spatially stratified sampling. All variables are then interpolated from irregular nodes to a uniform $H \times W$ grid using nearest-neighbor gridding, from which $\Delta x, \Delta y$ are computed. Time is converted to seconds using the NetCDF units, and Δt is set to the median positive time difference. Training and physics residuals are evaluated only on a wet-region mask M_{wet} (e.g., $h > 0.5$m and finite targets at $t = 0$). Finally, (u, v, ζ) are standardized using mean and standard deviation over wet pixels across all timesteps, and predictions are de-standardized before computing SWE residuals and reporting metrics.

b) Surrogate Formulation: In addition to explicit SWE time stepping, we include a proof-of-concept physics-informed neural surrogate to show that the same pipeline can be driven by a learned model.

Surrogate objective: The surrogate learns a mapping from terrain or bathymetry input to spatiotemporal fields:

$$G : B \to \{u_t, v_t, \eta_t\}_{t=1}^{T}, \tag{15}$$

implemented as a convolutional model that predicts the full time sequence in one forward pass. Training uses data loss on available targets (u, v, η) and physics loss that penalizes SWE residuals computed by finite differences across time and space:

$$L = L_{\mathrm{data}} + \lambda \left(L_{\mathrm{cont}} + L_{\mathrm{mom},u} + L_{\mathrm{mom},v}\right). \tag{16}$$

Here, L_{data} is the supervised reconstruction loss, L_{cont} is the continuity residual penalty, and $L_{\mathrm{mom},u}$ and $L_{\mathrm{mom},v}$ are the SWE momentum residual penalties in the x and y directions. The scalar $\lambda \ge 0$ balances data fitting and physics residuals; in our experiments, we use $\lambda = 1.0$ for stable training.

In rendering, the surrogate outputs are time-interpolated to match video FPS, used to advect a persistent water state, and combined with road and vehicle constraints from Module A. This enables fast video-scale inference while keeping the model grounded through physics residuals during training.

D. Photorealistic Rendering & Annotation

The final stage composites the simulated fluid state

$$\Phi_t = \{h, u, v\} \tag{17}$$

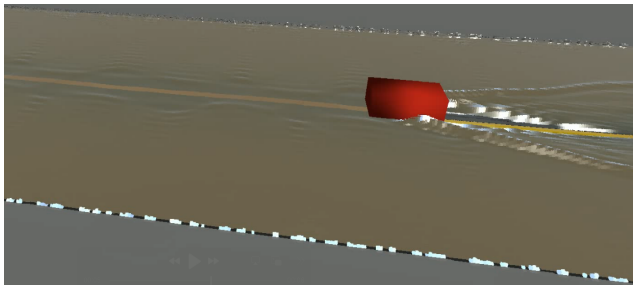

Fig. 2. Toy SWE validation: wake formation and drainage on a crowned road.

onto the original RGB frame I_t using a visualization model grounded in fluid optics.

1) Micro-Ripple Shading: To mimic surface tension and wavelets, we compute a shading layer derived from the surface gradients of the fluid height. We approximate the Fresnel effect by modulating intensity based on the local surface normal:

$$\mathcal{I}_{\mathrm{shade}} = \nabla(G_\sigma * h) \cdot L, \tag{18}$$

where G_σ is a Gaussian smoothing kernel and L is the incident light vector. This produces specular highlights on wave crests and shadows in troughs.

2) Automated Ground Truth Generation: Because the flood is generated via simulation, the dataset is fully annotated by construction. For every frame, we extract pixel-perfect labels for:

- **Semantic Segmentation:** High-fidelity masks for roads M_{road} and vehicles M_{veh}.
- **Flood Extent:** A binary mask defined strictly where the simulated water height $h > 0$.
- **Metric Flood Depth:** The precise per-pixel depth values $h(x, y)$ in meters, derived directly from the physics engine.

IV. RESULTS AND DISCUSSION

In this section, we qualitatively evaluate the proposed framework. We first validate the differentiable SWE solver on a controlled toy problem to establish numerical plausibility, then present results from the full pipeline on real-world monocular street videos. Finally, we report proof-of-concept results for the physics-informed neural surrogate described in Section C.2 and discuss trade-offs relative to GAN-based rendering (e.g., SURFGenerator [7]) and 3D Digital Twins [15].

A. Validation on Canonical "Toy Problem"

To verify that our implementation of the Shallow Water Equations (SWE) correctly models urban fluid dynamics, we constructed a simplified synthetic test case on a grid with two key geometric features: (i) a **road crown**, modeled as a parabolic surface elevation peaking at the center (z=0.35m) and tapering to the edges (z=0m) to mimic standard street drainage profiles, and (ii) a **dynamic obstacle**, modeled as a rectangular rigid body (2m $\times$ 4.5m) moving longitudinally at 10 m/s to represent a vehicle.

Observations: The simulation results in Figure 2 demonstrate two critical physical behaviors. First, as the obstacle

Fig. 3. Visualization of the Scene Decomposition module. Detected bounding boxes identify vehicles and guide the generation of precise segmentation masks, defining rigid obstacles used in our physics-based simulation.

moves through fluid at rest, a bow wave forms at the leading edge and a low-pressure wake forms at the trailing edge. The reflective boundary condition

$$u \cdot n = 0 \tag{19}$$

forces flow separation around the "vehicle." Second, when rain source terms are added, water naturally accumulates in the "gutters" at the grid edges rather than the center, confirming that the topographic gradient terms (∇B) correctly drive gravitational flow. These results show that the solver reproduces essential urban hydraulic phenomena, including drainage and vehicle-induced turbulence.

B. Qualitative Evaluation of Scene Decomposition

We applied the scene decomposition module to diverse monocular dashcam videos. Figure 3 shows intermediate outputs of the detection-segmentation pipeline. As shown, Grounding DINO successfully localizes vehicles with tight bounding boxes, filtering out background clutter. These boxes then serve as prompts for SAM 2, which generates precise pixel-level masks shown as colored overlays.

The integration of SAM 2 (Large) yielded robust temporal tracking. Unlike frame-by-frame segmentation methods that often exhibit flickering edges, SAM 2 maintained consistent vehicle boundaries across frames. By averaging road predictions over multiple frames, the system filtered out transient occlusions and reconstructed a clean, contiguous drivable surface even when the road was 60% occluded by cars. Depth Anything V2 provided the relative slope information needed to preserve the sharp curb-road gradient, preventing floodwater from leaking onto sidewalks prematurely.

C. Flood Simulation Results

Figure 4 illustrates the final output of the pipeline. The simulated floodwaters exhibit realistic interactions with the environment that purely generative models struggle to capture, including **wake formation** as vehicles move through water and generate directional ripples that interact with the road curb, **depth-dependent opacity** where shallow water near the road crown appears semi-transparent and deeper water in the gutters appears opaque and turbid, and **occlusion handling** in which alpha blending respects the vehicle masks from Figure 3 so water appears around the tires rather than over them. Most importantly, the generated data includes dense, pixel-perfect

Fig. 4. Final flood simulation. The simulation generates physically consistent results, displaying wake formation behind vehicles and realistic, depth-dependent water opacity.

TABLE I
EVALUATION OF TRAINED SURROGATE MODEL.

Metric	Overall	(u)	(v)	(ζ)
MSE	4.69528×10^{-6}	4.56440×10^{-6}	3.01270×10^{-6}	6.50873×10^{-6}
MAE	1.09071×10^{-3}	9.95074×10^{-4}	7.96090×10^{-4}	1.48098×10^{-3}
RMSE	2.16686×10^{-3}	2.13645×10^{-3}	1.73571×10^{-3}	2.55122×10^{-3}

depth annotations h, so water level is not merely a visual texture but a measurable physical quantity (in meters) at every pixel, governed by conservation of mass.

D. Physics-Informed Neural Surrogate

As a feasibility study, we trained a physics-informed neural surrogate (Section C.2) to predict (u, v, ζ) sequences from static bathymetry on a regular grid, regularized by SWE residual terms. The surrogate is not intended to replace the explicit solver; rather, it shows that physics-regularized predictions can provide coherent motion fields for rendering and be evaluated quantitatively on forecast data.

Table I reports wet-region masked metrics in physical units for the trained surrogate.

Synthetic flooding rendering: When integrated into the rendering stage, surrogate-predicted velocity fields provide temporally coherent advection cues, producing plausible flood progression under the same road and vehicle constraints used by the solver-driven pipeline. We emphasize that this surrogate is a proof of concept meant to illustrate the viability of physics-informed neural fields for driving motion, not equivalence to the explicit SWE solver for all urban scenes.

E. Qualitative Comparison

We compare our framework to two dominant paradigms for synthetic flood generation: purely generative models and high-fidelity 3D digital twins. Generative methods can produce realistic textures but often violate physics due to limited 3D understanding, for example by placing water on vertical surfaces or producing flow that ignores gravity. Digital twins offer strong physical accuracy, but rely on manually built 3D assets and expensive offline simulation and rendering, making them difficult to scale.

Fig. 5. Physics solver (left) and surrogate model (right).

Our approach bridges these trade-offs by operating directly on raw monocular video to preserve photorealism while enforcing physically consistent dynamics with the Shallow Water Equations (SWE), enabling scalable synthesis of annotated flood data with dense depth labels.

V. LIMITATIONS AND FUTURE DIRECTIONS

Limitations: The method is designed for monocular video, so the recovered scene geometry is relative and requires lightweight calibration or priors (e.g., road crown) for stable SWE simulation. Realism and stability also depend on temporally consistent road and vehicle masks, making robust segmentation important. In addition, SWE is a depth-averaged model well suited for scalable urban flow synthesis, while fine 3D spray effects are handled at the rendering layer rather than by the solver.

Future directions: This work demonstrates a proof-of-concept physics-informed surrogate model; the next step is to develop it into a solver-grade accelerator for the SWE module by conditioning on richer forcings (e.g., rain and obstacles) and strengthening physics and stability regularization. In parallel, we will reduce current sensitivities by adding lightweight metric calibration cues to address monocular depth scale ambiguity and improving mask robustness through temporal aggregation and confidence-aware filtering.

VI. CONCLUSION

This work presents a framework for generating physically grounded synthetic videos of urban flooding. We use Grounding DINO and SAM 2 for spatiotemporal scene decomposition and a differentiable SWE-based physics engine to simulate fluid dynamics directly on estimated terrain. The results show that the framework bridges visual realism and physical validity, producing realistic behaviors such as wake formation and drainage. We also provide a proof-of-concept PIN surrogate model that predicts (u, v, η) under SWE residual regularization, highlighting a promising direction for accelerating simulation and scaling dataset generation. We are expanding this pipeline to generate a fully annotated synthetic dataset with depth ground truth for quantitative risk analysis by leveraging ubiquitous RGB cameras rather than LiDAR or depth sensors.

ACKNOWLEDGMENTS

This research was partially supported by National Science Foundation grants 100876-010 and 1828593 at Old Dominion University. The authors acknowledge the use of Grounding DINO and SAM2, and thank Dr. Mecit Cetin (Old Dominion University) and Jonathan L. Goodall (University of Virginia) for providing data.

REFERENCES

[1] W. Kron, "Flood risk = hazard • values • vulnerability," *Water International*, vol. 30, no. 1, pp. 58–68, 2005, doi: 10.1080/02508060508691837.

[2] J. Sanyal and X. X. Lu, "Application of Remote Sensing in Flood Management with Special Reference to Monsoon Asia: A Review," *Natural Hazards*, vol. 33, no. 2, pp. 283–301, 2004, doi: 10.1023/B:NHAZ.0000037035.65105.95.

[3] S. W. Lo, J. H. Wu, F. P. Lin, and C. H. Hsu, "Visual Sensing for Urban Flood Monitoring," *Sensors*, vol. 15, no. 8, pp. 20006–20029, 2015, doi: 10.3390/s150820006.

[4] M. Rahnemoonfar *et al.*, "FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding," *IEEE Access*, vol. 9, pp. 89644–89654, 2021, doi: 10.1109/ACCESS.2021.3090981.

[5] D. Bonafilia, B. Tellman, and T. Anderson, "Sen1Floods11: A Georeferenced Dataset to Train and Test Deep Learning Flood Algorithms for Sentinel-1," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, 2020, pp. 835–845, doi: 10.1109/CVPRW50498.2020.00113.

[6] P. Zhong, Y. Liu, H. Zheng, and J. Zhao, "Detection of urban flood inundation from traffic images using deep learning methods," *Water Resources Management*, vol. 38, no. 1, pp. 287–301, 2024.

[7] S. Lamczyk, K. Ampofo, B. Salashour, M. Cetin, and K. M. Iftekharuddin, "SURFGenerator: Generative Adversarial Network Modeling for Synthetic Flooding Video Generation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, 2022, pp. 1–8, doi: 10.1109/IJCNN55064.2022.9891969.

[8] C. Sazara, M. Cetin, and K. M. Iftekharuddin, "Detecting floodwater on roadways from image data with handcrafted features and deep transfer learning," arXiv preprint arXiv:1909.00125, 2019, doi: 10.48550/arXiv.1909.00125.

[9] P. Chaudhary, S. D'Aronco, J. P. Leitão, K. Schindler, and J. D. Wegner, "Water level prediction from social media images with a multi-task ranking approach," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 167, pp. 252–262, 2020, doi: 10.1016/j.isprsjprs.2020.07.003.

[10] M. Witherow, C. Sazara, I. Winter-Arboleda, M. Elbakary, M. Cetin, and K. M. Iftekharuddin, "Floodwater detection on roadways from crowdsourced images," *Comput. Methods Biomech. Biomed. Eng. Imag. Vis.*, vol. 7, 2018, doi: 10.1080/21681163.2018.1488223.

[11] P. Chaudhary, S. D'Aronco, M. Moy de Vitry, J. P. Leitão, and J. D. Wegner, "Flood-water level estimation from social media images," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. IV-2/W5, 2019, doi: 10.5194/isprs-annals-IV-2-W5-5-2019.

[12] V. Schmidt *et al.*, "ClimateGAN: Raising Climate Change Awareness by Generating Images of Floods," arXiv:2110.02871, 2021.

[13] A. Luccioni, V. Schmidt, V. Vardanyan, Y. Bengio, and T. M. Rhyne, "Using Artificial Intelligence to Visualize the Impacts of Climate Change," *IEEE Computer Graphics and Applications*, vol. 41, no. 1, pp. 8–14, Jan./Feb. 2021, doi: 10.1109/MCG.2020.3025425.

[14] V. Schmidt *et al.*, "Visualizing the Consequences of Climate Change Using Cycle-Consistent Adversarial Networks," arXiv:1905.03709, 2019.

[15] M. M. Rahman, M. Umair, K. Ampofo, A. Temtam, and K. M. Iftekharuddin, "3D digital twin reconstruction of street flooding," in *Opt. Photon. Inf. Process. XIX*, 2025, p. 23, doi: 10.1117/12.3064661.

[16] S. Liu *et al.*, "Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection," arXiv:2303.05499, 2024.

[17] N. Ravi *et al.*, "SAM 2: Segment Anything in Images and Videos," arXiv:2408.00714, 2024.

[18] L. Yang *et al.*, "Depth Anything V2," arXiv:2406.09414, 2024.

[19] NOAA. Great Lakes Coastal Forecasting System (GLCFS) — Lake Superior forecast NetCDF subset (THREDDS). Great Lakes Environmental Research Laboratory (GLERL).