Supplementary Materials: ANFluid: Animate Natural Fluid Photos base on Physics-Aware Simulation and Dual-Flow Texture Learning

Anonymous Authors

1 USER STUDY

1.1 Questionnaire Production

The user study aims to evaluate whether our method indeed provides perceptible improvements compared to other methods when used by users. To ensure that our scene perception truly comes into play, we selected four scenarios from the CLAW dataset, comprising two scenarios suitable for SWE and PIC each. The specific scenario images are as follows Figure. 1:

Figure 1: Samples.Four scenarios from the CLAW dataset, comprising two scenarios suitable for SWE and PIC each.

We provided participants with masked optical flow images to input, along with fluid animation results synthesized using the mentioned three methods(SLR, Runway, Ours) in article, and asked users to rank the effects of a certain dimension from high to low across different scenarios.

To visually present the performance of each method in various aspects, we designed five evaluation metrics: video quality, picture integrity, fluid motion authenticity, editing comprehension ability, and texture realism. Here are the detailed explanation of each item.

- Video quality: To assess the The detail of the video.
- Picture integrity: To assess whether there are any occurrences of gaps in the generated results as motion occurs.
- Fluid motion authenticity: To evaluate the physical realism of the fluid component.
- Editing comprehension ability: To evaluate whether the generated results align with user expectations during guidance.
- Texture realism: To assess the degree of texture preservation compared to the original scene.

1.2 Questionnaire Collection

In the end, we collected a total of 170 questionnaires from China, the detailed distribution showed in Table 1 and Table 2.

Table 1: Survey details

Indicators count	Number of Questions
5	5

Table 2: Collection details

Number of Cities	Area Span	Personnel Count
11	Latitude:22.54-41.79 Longitude:114.08-123.42	170

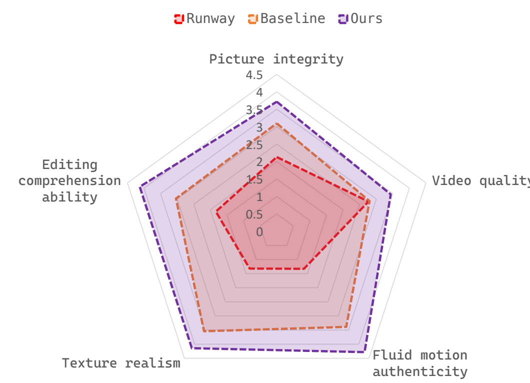


Figure 2: User study. The subjective ratings are derived from a subset of samples from the CLAW dataset, covering a balanced representation of both PIC and SWE scenes.

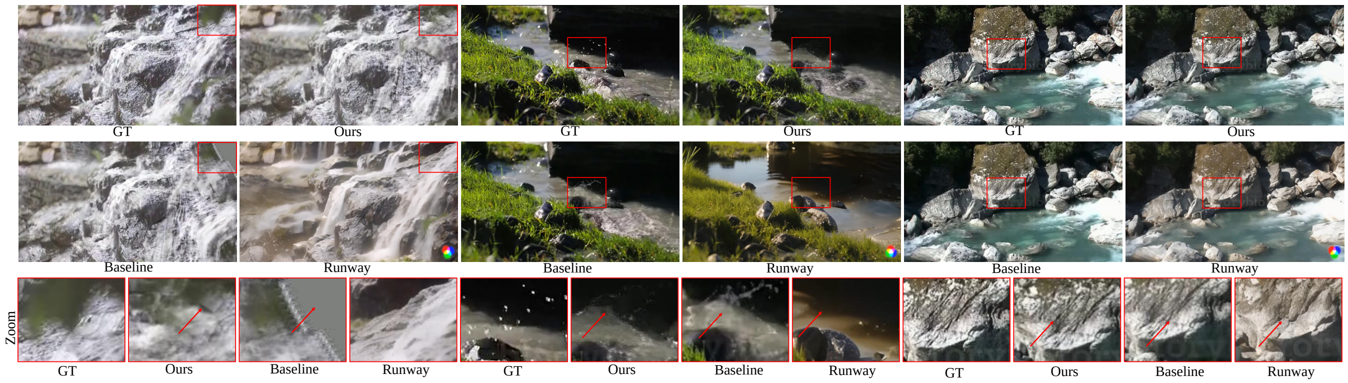
1.3 Data Processing

By aggregating all the surveys, we obtained the rankings (1/2/3) frequencies of each method across every dimension. To provide a more intuitive representation of the strengths and weaknesses of different methods across various dimensions, we employ the following formula to transform the ranking data into a parameter indicator (Eq. 1):

$$S_{\text{method,item}} = \frac{\alpha \times \text{first\_num} + \beta \times \text{second\_num} + \gamma \times \text{third\_num}}{\text{sample\_num}} \quad (1)$$

$$\text{sample\_num} = \text{first\_num} + \text{second\_num} + \text{third\_num} \quad (2)$$

Where $S_{\text{method,item}}$ represents the final average score of the results obtained using a particular method on a specific dimension. first_num, second_num, and third_num respectively denote the number of times users selected it as first, second, and third choices. sample_num represents the total number of samples. Hence, there is a constraint formula 2. α, β, γ denote the weighting factors for the rankings. In this statistical analysis, we set $\alpha = 5$, $\beta = 3$, $\gamma = 1$.



(a) Hole (b) Texture (c) Blurriness

Figure 3: Comparison of Effects. More comparative analysis demonstrates the effectiveness of our approach.

1.4 Study Result

The final results are presented in Figure 2.

Different colors represent different methods, where the better the performance of a specific metric, the further outward the corresponding branch extends. Our approach has demonstrated multifaceted enhancements. Notably, there has been a significant boost in the perceived credibility of fluid motion and overall image coherence over the baseline, underscoring the efficacy of the PAS and DFTL techniques. Additionally, unexpected improvements were also observed in video quality and editing comprehension.

1.5 More Comparative Effects

As illustrated in Figure 3, our DFTL network leverages bidirectional self-supervised optical flow estimation and multi-scale wrapping techniques to enhance dynamic relationships and elevate the overall quality of animations. This approach effectively addresses common issues encountered in animations such as holes, unrealistic textures, and blurriness.

By integrating bidirectional self-supervised optical flow estimation, we are able to accurately capture the motion and deformation of objects within the animation, thus maintaining continuity and realism more effectively. Additionally, the use of multi-scale wrapping enhances the clarity and realism of textures and details in the animation, eliminating blurriness and unrealistic appearances, thereby further enhancing the visual quality and perceptual appeal of the animations.