# Supplementary Materials for "Fixing Malfunctional Objects With Learned Physical Simulation and Functional Prediction"

- In Section A, we provide more details on the dataset.
- In Section **B**, we provide more details on the experimental setup
- In Section C, we show more experimental results.

### A. Dataset

In Table 1, we show the fixing types (scale, translate, rotate) that can be applied to each category and included in the choices. In Figure 1, we add more examples of our dataset.

Category	Refrigerator	Bucket	USB	Kettle	Cart	KitchenPot	Box
Fixing Type	S, T, R	S, T	S, T	R	R	R	S

Table 1. The fixing types that can be applied to the categories. S=Scale, T=Translate, R=Rotate



Figure 1. Examples of our FIXIT dataset.



Figure 2. Qualitative Examples of FixNet

## **B.** Implementation Details

#### **B.1. DSL-based baselines**

For the DSL-based baselines, we input the "type" and "axis" into an embedding layer and then pass through a LSTM layer to extract language features. The value is fed into a linear layer and concatenated with the language features. The DSL features go through MLP to output binary classification results.

#### **B.2.** FixNet

For the flow proposal network, we pretrain FlowNet3D on the Flyingthings3D [1] dataset, and finetuned on 10% of the ground-truth flows of the videos in our training set. We use a learning rate of 0.001, batch size of 32, and dropout of 0.5. We train the model for 250 epochs, and use the model that achieves best performance on the validation dataset.

For the instance segmentation network, we use a learning rate of 0.001, decay rate of 1e-4 for weights, decay steps for learning rate as 20 and decay rate of 0.5 for learning rate. The batch size is 16. We use Adam Optimizer to train for 250 epochs.

The physical dynamics prediction module is trained for 1.5M iterations till converge using Adam optimizer with initial learning rate 1e-4.

For the funcitonality prediction network, we use learning rate of 0.001 and decay rate of 1e-4 to train for 200 epochs. The batch size is 24.

#### **C. More Qualitative Examples**

In Figure 2, we show some more qualitative examples output by FixNet.

#### References

 Nikolaus Mayer, Eddy Ilg, Philip Hausser, Philipp Fischer, Daniel Cremers, Alexey Dosovitskiy, and Thomas Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In *Proceedings of the IEEE conference on computer* vision and pattern recognition, pages 4040–4048, 2016. 2