

Context-Memorized Dataset Distillation for Length-Robust Sensor-Grounded Summarization

Jason K. Nam^{1,2}

1 Department of Data Science, Hongik University 2 Artificial Intelligence Research Division, ETRI
skarlgja68@gmail.com

Shared report supervision across nested contexts.

Serialized surveillance reports. Frozen 3B and 7B backbones. LoRA/QLoRA adapter tuning.

01 Input formats

O2O

One object row / object phrase

M2O

Frame-table prefix / frame report

M2M

Multiple frames / episode report

Nested contexts repeat identical or overlapping report tokens.

02 Training schedule



One adapter continues along a selected path.
Shared targets carry knowledge across lengths without an external teacher or synthetic examples.

Optional cosine gradient matching aligns paired tokens on heterogeneous edges.

03 Cross-format gap



O2O-to-M2O has the smallest Val1/Val3 semantic gap among the focused paths.

Gap = |SemAvg(Val1) - SemAvg(Val3)|
Lower means less cross-format variation.

Canonical M2O results

Model	Training path	Repeats	SemAvg	PPL
3B	LaKD	5	0.497	2.860
3B	O2O-to-M2O	2	0.484	1.036
7B	LaKD	5	0.485	3.168
7B	O2O-to-M2O	2	0.493	1.035

SemAvg = (ROUGE-L + BERTScore) / 2. Higher is better. PPL: lower is better.

The best path depends on the target

O2O-to-M2O: stable M2O transfer and low PPL.

LaKD: strongest canonical SemAvg for 3B.

M2O-to-M2M: specialization for temporal reports.

Reference-based results from the reported runs.