MVP-N: A Dataset and Benchmark for Real-World Multi-View Object Classification

Ren Wang[1], Jiayue Wang[1], Tae Sung Kim[2], Jin-Sung Kim[2], Hyuk-Jae Lee[1]
[1]Seoul National University [2]Sun Moon University

2022 NeurIPS Datasets and Benchmarks Track

Ren Wang

Overview

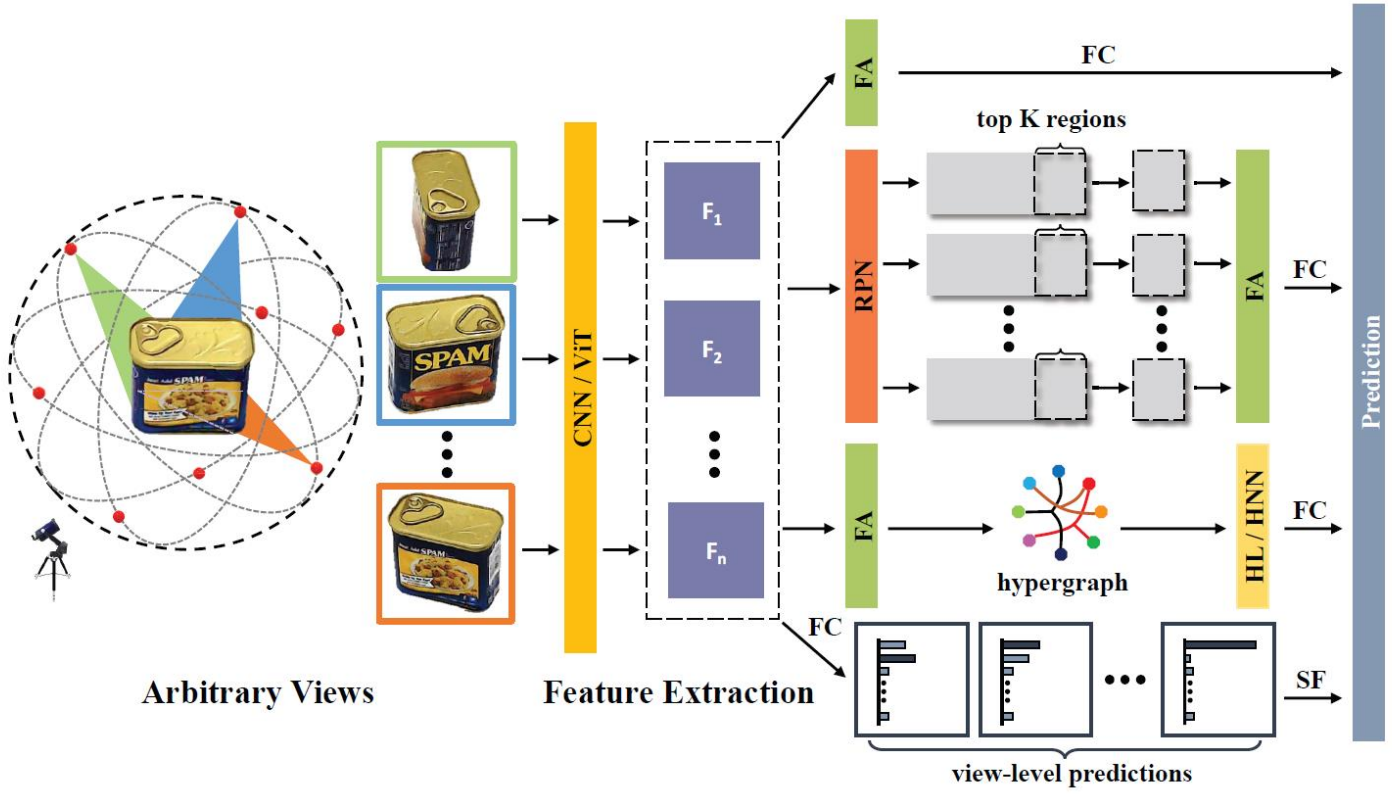
- Why multi-view?
  - Human visual perception of 3D objects relies on 2D observations from different perspectives.
  - Single-view representations may not provide discriminative features.

- Contributions
  - Construct a real-world fine-grained dataset with HPIQ annotations for multi-view object classification.
  - Benchmark 4 multi-view-based feature aggregation methods and 12 soft label methods on MVP-N.
  - Propose a new metric and an evaluation protocol based on HPIQ annotations for soft label methods.

https://github.com/SMNUResearch/MVP-N

Task: Multi-View Object Classification

- Existing methods
  - Two-stage
    - Feature aggregation
    - Score fusion
  - Three-stage
    - Hypergraph-based
    - Part-based



- Properties of practical methods
  - Arbitrary number of views
  - Free view configurations
  - Unknown camera positions and relative poses

Motivation

- Limitations of existing datasets
  - Synthetic polygon mesh objects
  - Coarse-grained categorization
  - No validation split
  - Lack of view-level annotations

	RGB-D Object	ModelNet40	MIRO	ScanObjectNN	FG3D	**MVP-N (ours)**
Year	2011	2015	2018	2019	2021	2022
Representation	RGB-D	Mesh	RGB	Point Cloud	Mesh	RGB
#Categories	51	40	12	15	66	44
Real-world objects	✓	✗	✓	✓	✗	✓
Real capture environment	✓	✗	✓	✗	✗	✓
Fine-grained	✗	✗	✗	✗	✓	✓
Validation set	✗	✗	✗	✗	✗	✓
View-level annotation	✗	✗	✗	✗	✗	✓

MVP-N: Design and Construction

- Step 1: Object selection
  - 44 fine-grained retail products
  - High inter-class view similarity (multi-view noise)
- Step 2: Data collection
- Step 3: Data annotation
  - Information quantity judgment
  - Bounding box annotation
- Step 4: Quality control and data filtering
  - ‘Informative/Uninformative’ (HPIQ) annotation
- Step 5: Data preprocessing
- Step 6: Train/valid/test split
  - View sampling: 16k
  - Multi-view set construction: 9k
    - 2 to 6 views

Benchmark on MVP-N

- Multi-view-based feature aggregation
  - 4 methods
  - Evaluation metric
    - Multi-view accuracy (MVA)
    - Mean confidence for correct predictions (MCC)
    - Mean confidence for wrong predictions (MCW)
    - Model size
    - Number of floating-point operations (FLOPs)
    - Inference latency
- Soft label
  - 12 methods
  - Evaluation metric
    - Multi-view accuracy (MVA)
    - Single-view accuracy (SVAI)
    - Mean confidence for correct predictions (MCCI)
    - Mean confidence for wrong predictions (MCWI)
    - Mean confidence difference between predictions and ground truths (MCDU)

Results: Multi-View-Base Feature Aggregation

Method	MVA (%) ↑	MCC ↑	MCW ↓	Model Size (M) ↓	FLOPs (G) ↓	Latency (ms) ↓
Validation:						
MVCNN-new	89.29 ± 0.88	$\mathbf{0.8812 \pm 0.0040}$	0.6568 ± 0.0120	**11.20**	**10.91**	$\mathbf{6.23 \pm 0.03}$
GVCNN	85.69 ± 1.01	0.8275 ± 0.0044	$\mathbf{0.6095 \pm 0.0136}$	24.04	10.99	7.60 ± 0.07
DAN	$\mathbf{92.05 \pm 0.56}$	0.8592 ± 0.0044	0.6192 ± 0.0055	17.50	10.95	8.11 ± 0.04
CVR	79.95 ± 1.89	0.8347 ± 0.0118	0.6564 ± 0.0157	34.38	11.08	12.57 ± 0.07
Test:						
MVCNN-new	89.35 ± 1.21	$\mathbf{0.8792 \pm 0.0053}$	0.6552 ± 0.0069	**11.20**	**10.91**	$\mathbf{6.23 \pm 0.03}$
GVCNN	85.42 ± 1.37	0.8267 ± 0.0032	$\mathbf{0.6055 \pm 0.0088}$	24.04	10.99	7.60 ± 0.07
DAN	$\mathbf{91.61 \pm 0.94}$	0.8602 ± 0.0050	0.6211 ± 0.0062	17.50	10.95	8.11 ± 0.04
CVR	79.99 ± 2.52	0.8339 ± 0.0127	0.6457 ± 0.0166	34.38	11.08	12.57 ± 0.07

Results: Soft Label

Method	SVA (%)	SVAI (%) ↑	MCCI ↑	MCWI ↓	MCDU ↓	MVA (%) ↑
Validation:						
CE	76.76 ± 0.24	99.44 ± 0.17	0.9475 ± 0.0031	0.6076 ± 0.0368	0.3977 ± 0.0091	83.05 ± 0.56
KD	78.47 ± 0.55	99.62 ± 0.08	0.9587 ± 0.0009	0.5799 ± 0.0295	0.3867 ± 0.0040	85.72 ± 1.24
SB	74.41 ± 0.36	99.08 ± 0.33	0.8911 ± 0.0074	0.5573 ± 0.0177	0.2945 ± 0.0046	83.31 ± 0.41
HB	76.69 ± 0.18	99.44 ± 0.16	0.9469 ± 0.0029	0.6073 ± 0.0369	0.3989 ± 0.0097	82.73 ± 0.60
LS	76.03 ± 0.36	99.26 ± 0.15	0.7711 ± 0.0056	$\mathbf{0.4101 \pm 0.0262}$	0.2534 ± 0.0093	84.30 ± 1.05
DSB	76.06 ± 0.98	99.15 ± 0.58	0.9148 ± 0.0522	0.5704 ± 0.0313	0.3577 ± 0.0626	82.71 ± 0.69
DHB	76.67 ± 0.27	99.48 ± 0.18	0.9454 ± 0.0022	0.6113 ± 0.0301	0.3971 ± 0.0069	82.60 ± 0.70
SAT	74.55 ± 0.40	99.18 ± 0.19	0.8746 ± 0.0049	0.5465 ± 0.0179	0.2256 ± 0.0058	86.52 ± 0.36
LRT	76.57 ± 0.52	99.60 ± 0.15	$\mathbf{0.9609 \pm 0.0018}$	0.6094 ± 0.0642	0.4240 ± 0.0104	84.29 ± 1.26
SEAL	71.97 ± 0.33	98.92 ± 0.23	0.6846 ± 0.0036	0.4379 ± 0.0102	$\mathbf{0.1404 \pm 0.0018}$	85.48 ± 0.65
PLC	76.51 ± 0.27	99.33 ± 0.20	0.9469 ± 0.0033	0.6126 ± 0.0424	0.4042 ± 0.0119	82.37 ± 0.72
OLS	76.63 ± 0.14	99.30 ± 0.17	0.9336 ± 0.0041	0.5852 ± 0.0273	0.3774 ± 0.0101	82.90 ± 0.57
HPIQ	62.77 ± 0.42	$\mathbf{99.73 \pm 0.04}$	0.9246 ± 0.0057	0.5538 ± 0.0447	0.1530 ± 0.0068	$\mathbf{93.55 \pm 0.79}$
Test:						
CE	78.65 ± 0.44	99.15 ± 0.11	0.9383 ± 0.0028	0.6035 ± 0.0442	0.3892 ± 0.0070	83.37 ± 1.05
KD	80.38 ± 0.24	99.49 ± 0.09	0.9509 ± 0.0014	0.5574 ± 0.0606	0.3737 ± 0.0014	86.77 ± 1.24
SB	76.22 ± 0.27	98.73 ± 0.15	0.8789 ± 0.0068	0.5230 ± 0.0210	0.2862 ± 0.0107	83.85 ± 0.84
HB	78.52 ± 0.51	99.10 ± 0.15	0.9376 ± 0.0022	0.6050 ± 0.0305	0.3899 ± 0.0068	83.20 ± 1.11
LS	77.65 ± 0.30	98.82 ± 0.30	0.7522 ± 0.0054	$\mathbf{0.3843 \pm 0.0309}$	0.2474 ± 0.0114	83.96 ± 1.50
DSB	77.73 ± 0.81	98.90 ± 0.35	0.9037 ± 0.0523	0.5784 ± 0.0606	0.3457 ± 0.0610	83.09 ± 0.83
DHB	78.34 ± 0.46	99.07 ± 0.17	0.9365 ± 0.0023	0.6040 ± 0.0488	0.3871 ± 0.0041	83.07 ± 0.87
SAT	76.28 ± 0.43	99.00 ± 0.14	0.8620 ± 0.0049	0.5293 ± 0.0337	0.2145 ± 0.0063	87.37 ± 1.15
LRT	77.85 ± 0.46	99.33 ± 0.19	$\mathbf{0.9542 \pm 0.0013}$	0.5881 ± 0.0327	0.4076 ± 0.0141	83.78 ± 2.05
SEAL	73.58 ± 0.54	98.41 ± 0.25	0.6674 ± 0.0033	0.4018 ± 0.0085	$\mathbf{0.1326 \pm 0.0011}$	86.42 ± 0.74
PLC	78.40 ± 0.47	99.07 ± 0.11	0.9383 ± 0.0031	0.6070 ± 0.0370	0.3948 ± 0.0106	82.96 ± 1.06
OLS	78.40 ± 0.49	99.00 ± 0.12	0.9225 ± 0.0029	0.5799 ± 0.0239	0.3684 ± 0.0080	83.35 ± 1.05
HPIQ	63.31 ± 0.38	$\mathbf{99.68 \pm 0.10}$	0.9186 ± 0.0059	0.5934 ± 0.0222	0.1481 ± 0.0076	$\mathbf{94.36 \pm 0.56}$

Results: Influence of the number of uninformative views

Method	Validation					Test				
	2 views	3 views	4 views	5 views	6 views	2 views	3 views	4 views	5 views	6 views
feature aggregation:										
MVCNN-new	91.93 ± 0.77	88.80 ± 0.54	88.30 ± 1.17	88.25 ± 1.04	89.18 ± 1.52	89.52 ± 1.38	88.98 ± 1.31	87.36 ± 1.80	89.09 ± 1.76	91.82 ± 1.01
GVCNN	93.23 ± 1.14	84.73 ± 0.77	84.48 ± 1.52	81.93 ± 1.47	84.09 ± 1.40	89.70 ± 1.59	82.43 ± 1.92	84.05 ± 1.65	82.70 ± 1.87	88.20 ± 1.06
DAN	93.80 ± 0.92	91.59 ± 1.12	91.07 ± 0.99	$\mathbf{91.50 \pm 1.24}$	$\mathbf{92.32 \pm 0.63}$	91.48 ± 0.44	91.20 ± 1.17	89.59 ± 1.12	91.39 ± 1.51	$\mathbf{94.41 \pm 1.01}$
CVR	86.16 ± 0.73	81.16 ± 1.82	79.25 ± 2.55	76.39 ± 2.85	76.80 ± 2.51	84.20 ± 2.19	79.95 ± 2.25	77.80 ± 3.72	77.52 ± 2.94	80.45 ± 3.42
soft label:										
CE	92.98 ± 0.59	82.14 ± 0.97	80.25 ± 0.81	79.23 ± 1.16	80.64 ± 0.54	90.82 ± 1.42	79.95 ± 1.33	79.34 ± 1.66	80.61 ± 1.65	86.11 ± 0.52
KD	95.84 ± 0.47	84.48 ± 1.03	83.68 ± 1.99	81.39 ± 1.68	83.20 ± 1.71	94.64 ± 0.85	84.57 ± 2.49	82.50 ± 1.33	84.05 ± 1.31	88.09 ± 1.82
SB	92.23 ± 0.73	83.77 ± 0.32	80.55 ± 0.64	79.73 ± 1.04	80.30 ± 0.86	89.11 ± 1.26	82.57 ± 0.94	80.91 ± 1.26	81.09 ± 1.87	85.59 ± 0.60
HB	92.82 ± 0.88	81.73 ± 0.79	80.11 ± 1.11	78.59 ± 1.11	80.39 ± 0.64	90.80 ± 1.32	79.57 ± 1.25	79.50 ± 2.03	80.27 ± 1.84	85.86 ± 0.47
LS	90.84 ± 0.84	84.61 ± 1.35	82.84 ± 1.65	81.36 ± 0.76	81.84 ± 1.74	88.25 ± 1.32	82.09 ± 1.32	80.91 ± 1.78	81.64 ± 2.22	86.91 ± 1.02
DSB	92.34 ± 1.78	82.36 ± 0.94	79.93 ± 0.44	78.68 ± 1.11	80.25 ± 0.85	90.07 ± 1.28	80.39 ± 1.33	79.18 ± 1.39	79.93 ± 1.38	85.89 ± 0.85
DHB	92.95 ± 1.02	82.16 ± 0.53	79.91 ± 0.57	77.91 ± 1.36	80.09 ± 1.13	90.73 ± 1.40	79.98 ± 1.44	78.77 ± 1.31	80.39 ± 1.17	85.50 ± 0.65
SAT	94.27 ± 0.18	87.20 ± 0.68	84.39 ± 0.96	83.52 ± 1.12	83.23 ± 1.10	91.45 ± 1.21	87.25 ± 1.08	84.70 ± 1.76	84.93 ± 1.98	88.50 ± 0.86
LRT	94.93 ± 0.60	83.57 ± 1.25	81.77 ± 1.40	80.25 ± 1.93	80.91 ± 1.77	93.02 ± 1.19	81.07 ± 2.98	79.61 ± 1.97	80.45 ± 2.34	84.75 ± 2.21
SEAL	93.02 ± 1.08	86.64 ± 0.98	84.11 ± 0.60	81.61 ± 0.60	82.02 ± 0.85	90.80 ± 0.79	85.95 ± 0.96	83.75 ± 0.99	84.34 ± 0.79	87.27 ± 0.82
PLC	92.75 ± 0.54	81.45 ± 1.00	79.64 ± 0.70	78.39 ± 1.02	79.64 ± 1.21	90.61 ± 1.26	79.57 ± 1.06	79.09 ± 1.87	80.07 ± 1.68	85.48 ± 0.92
OLS	92.43 ± 0.56	81.91 ± 0.55	80.39 ± 1.13	79.18 ± 0.87	80.57 ± 0.71	90.00 ± 1.20	80.09 ± 1.30	79.66 ± 1.95	80.66 ± 1.41	86.32 ± 0.52
HPIQ	$\mathbf{98.34 \pm 0.36}$	$\mathbf{95.91 \pm 0.53}$	$\mathbf{93.23 \pm 1.15}$	90.59 ± 0.97	89.68 ± 1.88	$\mathbf{97.73 \pm 0.38}$	$\mathbf{96.36 \pm 0.47}$	$\mathbf{93.64 \pm 1.34}$	$\mathbf{92.02 \pm 0.86}$	92.07 ± 0.91