Method	RGB-D Object [3]		ModelNet40 [4]		ScanObjectNN [5]		FG3D [6] (Airplane)		FG3D [6] (Chair)		FG3D [6] (Car)	
	I	C	AIA	ACA	AIA	ACA	AIA	ACA	AIA	ACA	AIA	ACA
Two-stage:												
MVCNN [7]	-	-	90.1*	90.1*	-	-	91.11	82.57	82.90	76.27	76.12	71.88
Pairwise Network [8]	-	-	90.7	-	-	-	-	-	-	-	-	-
GIFT [9, 10]	-	-	89.50*	89.50*	-	-	-	-	-	-	-	-
RCPCNN [11]	-	-	92.2	-	-	-	-	-	-	-	-	-
GVCNN [12]	-	89.8‡	93.1*	93.1*	88.68	86.64	-	-	-	-	-	-
MHBN [13]	-	-	94.12	92.23	-	-	-	-	-	-	-	-
VERAM [14]	-	-	93.7	92.1	-	-	-	-	-	-	-	-
SeqViews2SeqLabels [15]	-	-	93.31	91.12	-	-	93.44	88.52	82.54	79.89	75.36	72.23
MVCNN-new [16]	-	89.0‡	95.9	93.4	87.82	85.71	-	-	-	-	-	-
MV-LSTM [17]	-	-	91.05	-	-	-	-	-	-	-	-	-
DeepCCFV [18]	-	-	92.46*	92.46*	-	-	-	-	-	-	-	-
MLVCNN [19]	-	-	94.16*	94.16*	-	-	-	-	-	-	-	-
3D2SeqViews [20]	-	-	93.40	91.51	-	-	92.21	89.41	77.10	76.26	66.39	63.85
RotationNet [1, 2]	97.45	99.51	97.37	96.29	86.90	84.88	92.76	89.11	82.07	78.45	75.59	72.53
3DViewGraph [21]	-	-	93.80*	93.80*	-	-	93.85	88.21	83.58	79.88	77.11	71.65
MVSG-DNN [22]	-	-	92.3	-	-	-	-	-	-	-	-	-
EMVN [23]	-	-	94.67	-	-	-	-	-	-	-	-	-
Relation Network [24]	-	-	94.3	92.3	-	-	-	-	-	-	-	-
View-GCN [25]	-	94.3	97.6	96.5	90.39	88.67	93.58	87.44	83.63	79.70	77.34	73.68
HEAR [26]	-	-	96.7	95.2	-	-	-	-	-	-	-	-
DRCNN [27]	-	-	96.84	94.86	-	-	-	-	-	-	-	-
VMM [28]	-	-	96.72	95.24	-	-	-	-	-	-	-	-
JointMVCNN [29]	-	-	94.16*	94.16*	-	-	-	-	-	-	-	-
MVLADN [30]	-	-	93.96	91.68	-	-	-	-	-	-	-	-
DRMV [31]	-	-	96.76	95.04	-	-	-	-	-	-	-	-
DAN [32]	-	-	93.5	-	-	-	-	-	-	-	-	-
CAR-Net [33]	-	-	97.73	-	-	-	-	-	-	-	-	-
CVR [34]	-	91.8‡	97.16	95.77	90.74	88.39	-	-	-	-	-	-
SVHAN [35]	-	-	94.3	-	-	-	-	-	-	-	-	-
MVT [36]	-	-	97.5	-	-	-	-	-	-	-	-	-
VFMVAC [37]	-	94.90	97.97	96.80	-	-	-	-	-	-	-	-
Three-stage (Hypergraph):												
iMHL [38]	-	-	97.16*	97.16*	-	-	-	-	-	-	-	-
HGNN [39]	-	-	96.7	-	-	-	-	-	-	-	-	-
MHGNN [40]	-	-	97.5	-	-	-	-	-	-	-	-	-
HGNN^{+} [41]	-	-	96.92	-	-	-	-	-	-	-	-	-
AMHCN [42]	-	-	97.86	-	-	-	-	-	-	-	-	-
GHSC [43]	-	-	97.75	-	-	-	-	-	-	-	-	-
Three-stage (Part):												
Parts4Feature [44]	-	-	93.40*	93.40*	-	-	91.39	82.55	81.61	77.08	75.44	73.42
FG3D-Net [6]	-	-	-	-	-	-	93.99	89.44	83.94	80.04	79.47	74.03

Table 1: **I**: Instance-level accuracy. **C**: Category-level accuracy. **AIA**: Average instance accuracy. **ACA**: Average class accuracy. '$-$': Results are not reported. '$\star$': The train/test split is 80 training samples (or all samples if the number is less than 80) and 20 test samples for each category. '‡': Results are quoted from [34]. Results on ScanObjectNN [5] and FG3D [6] are respectively quoted from [34] and [6]. All results are reported in percentage.

A Performance

Table 1 lists the performance of 39 multi-view-based feature aggregation methods on four existing datasets. Since MIRO [1] is only used in [1, 2], results are not included here.

Method	Configuration	MVA (%) ↑	MCC ↑	MCW ↓	Model Size (M) ↓	FLOPs (G) ↓	Latency (ms) ↓
DAN	$h = 1$	91.87 ± 0.90	0.8569 ± 0.0036	**0.6148 ± 0.0068**	**14.35**	**10.93**	**7.18 ± 0.07**
DAN	$h = 2$	**92.05 ± 0.56**	0.8592 ± 0.0044	0.6192 ± 0.0055	17.50	10.95	8.11 ± 0.04
DAN	$h = 3$	90.75 ± 1.18	0.8679 ± 0.0108	0.6340 ± 0.0124	20.66	10.97	8.98 ± 0.04
DAN	$h = 4$	90.65 ± 0.76	0.8852 ± 0.0073	0.6537 ± 0.0135	23.81	10.99	9.92 ± 0.03
DAN	$h = 5$	90.20 ± 1.33	0.8859 ± 0.0130	0.6619 ± 0.0137	26.96	11.01	10.83 ± 0.05
DAN	$h = 6$	89.54 ± 1.30	0.8977 ± 0.0087	0.6758 ± 0.0072	30.11	11.02	11.71 ± 0.05
DAN	$h = 7$	89.20 ± 0.86	**0.9015 ± 0.0043**	0.6913 ± 0.0111	33.27	11.04	12.62 ± 0.08
DAN	$h = 8$	88.87 ± 1.31	0.8990 ± 0.0074	0.6890 ± 0.0151	36.42	11.06	13.52 ± 0.06
CVR	$M = 4, \lambda = 0.1$	70.50 ± 6.61	0.8078 ± 0.0306	0.6432 ± 0.0518	**34.38**	**11.02**	**12.46 ± 0.09**
CVR	$M = 8, \lambda = 0.1$	64.18 ± 1.50	0.8175 ± 0.0058	0.6633 ± 0.0161	34.38	11.08	12.57 ± 0.07
CVR	$M = 16, \lambda = 0.1$	55.53 ± 3.07	0.8116 ± 0.0148	0.6589 ± 0.0260	34.39	11.18	12.79 ± 0.06
CVR	$M = 4, \lambda = 0.5$	77.44 ± 2.31	0.8075 ± 0.0181	0.6379 ± 0.0231	**34.38**	**11.02**	**12.46 ± 0.09**
CVR	$M = 8, \lambda = 0.5$	78.27 ± 2.04	0.8291 ± 0.0062	0.6484 ± 0.0116	34.38	11.08	12.57 ± 0.07
CVR	$M = 16, \lambda = 0.5$	76.93 ± 1.94	0.8271 ± 0.0106	0.6631 ± 0.0107	34.39	11.18	12.79 ± 0.06
CVR	$M = 4, \lambda = 1$	78.58 ± 1.41	0.8172 ± 0.0121	**0.6335 ± 0.0198**	**34.38**	**11.02**	**12.46 ± 0.09**
CVR	$M = 8, \lambda = 1$	**79.95 ± 1.89**	0.8347 ± 0.0118	0.6564 ± 0.0157	34.38	11.08	12.57 ± 0.07
CVR	$M = 16, \lambda = 1$	76.22 ± 2.38	**0.8358 ± 0.0130**	0.6808 ± 0.0174	34.39	11.18	12.79 ± 0.06

Table 2: Performance of multi-view-based feature aggregation methods with different configurations.

B Hyperparameter Search

B.1 Range of hyperparameters

DAN [32]: Number of sublayers $h \in [1, 2, 3, 4, 5, 6, 7, 8]$.

CVR [34]: Number of canonical views $M \in [4, 8, 16]$. Weighting factor $\lambda \in [0.1, 0.5, 1]$.

KD [45]: Temperature $T \in [2, 2.5, 3, 3.5, 4]$.

SB [46]: Weighting factor $\beta \in [0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95]$.

HB [46]: Weighting factor $\beta \in [0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95]$.

SEAL [47]: Number of iterations $I \in [1, 2, 3, 4, 5]$.

PLC [48]: Initial confidence threshold $\theta_0 \in [0.2, 0.3, 0.4, 0.5]$. Step size $\beta \in [0.05, 0.1, 0.2, 0.3]$.

OLS [49]: Weighting factor $\alpha \in [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9]$.

B.2 Results

Tables 2 and 3 show the results of hyperparameter search for two multi-view-based feature aggregation methods and six soft label methods, respectively. The best configuration for each method is marked with blue.

Method	Configuration	SVA (%)	SVAI (%) ↑	MCCI ↑	MCWI ↓	MCDU ↓	MVA (%) ↑
KD	$T = 2$	78.47 ± 0.55	**99.62 ± 0.08**	**0.9587 ± 0.0009**	0.5799 ± 0.0295	0.3867 ± 0.0040	**85.72 ± 1.24**
KD	$T = 2.5$	78.20 ± 0.69	99.58 ± 0.09	0.9584 ± 0.0012	0.5908 ± 0.0422	0.3837 ± 0.0046	85.58 ± 1.63
KD	$T = 3$	77.64 ± 0.76	99.59 ± 0.10	0.9569 ± 0.0015	0.5760 ± 0.0382	0.3827 ± 0.0050	84.99 ± 1.34
KD	$T = 3.5$	77.23 ± 0.59	99.52 ± 0.10	0.9546 ± 0.0012	**0.5751 ± 0.0398**	0.3817 ± 0.0061	84.41 ± 1.41
KD	$T = 4$	76.70 ± 0.75	99.49 ± 0.12	0.9525 ± 0.0014	0.5862 ± 0.0299	**0.3755 ± 0.0048**	83.50 ± 1.27
SB	$\beta = 0.5$	74.41 ± 0.36	99.08 ± 0.33	0.8911 ± 0.0074	**0.5573 ± 0.0177**	**0.2945 ± 0.0046**	**83.31 ± 0.41**
SB	$\beta = 0.55$	74.78 ± 0.31	99.13 ± 0.30	0.9040 ± 0.0045	0.5635 ± 0.0197	0.3193 ± 0.0100	82.88 ± 0.72
SB	$\beta = 0.6$	75.09 ± 0.36	99.16 ± 0.24	0.9112 ± 0.0062	0.5681 ± 0.0238	0.3320 ± 0.0116	82.59 ± 0.83
SB	$\beta = 0.65$	75.35 ± 0.35	99.24 ± 0.23	0.9175 ± 0.0056	0.5771 ± 0.0253	0.3440 ± 0.0105	82.66 ± 0.66
SB	$\beta = 0.7$	75.63 ± 0.27	99.25 ± 0.21	0.9209 ± 0.0053	0.5877 ± 0.0255	0.3492 ± 0.0138	82.88 ± 1.13
SB	$\beta = 0.75$	75.79 ± 0.32	99.28 ± 0.22	0.9296 ± 0.0033	0.5897 ± 0.0316	0.3686 ± 0.0097	82.31 ± 0.70
SB	$\beta = 0.8$	76.05 ± 0.30	99.34 ± 0.18	0.9340 ± 0.0027	0.5976 ± 0.0375	0.3772 ± 0.0099	82.48 ± 0.66
SB	$\beta = 0.85$	76.20 ± 0.25	99.33 ± 0.16	0.9384 ± 0.0027	0.5898 ± 0.0397	0.3850 ± 0.0128	82.46 ± 0.71
SB	$\beta = 0.9$	76.43 ± 0.21	99.40 ± 0.22	0.9412 ± 0.0037	0.6086 ± 0.0390	0.3879 ± 0.0102	82.66 ± 0.58
SB	$\beta = 0.95$	76.58 ± 0.17	**99.44 ± 0.17**	**0.9441 ± 0.0033**	0.6075 ± 0.0383	0.3913 ± 0.0095	82.77 ± 0.76
HB	$\beta = 0.5$	75.87 ± 0.48	99.32 ± 0.18	0.9407 ± 0.0029	**0.6039 ± 0.0287**	**0.3863 ± 0.0156**	82.44 ± 1.14
HB	$\beta = 0.55$	76.10 ± 0.42	99.40 ± 0.23	0.9421 ± 0.0046	0.6063 ± 0.0161	0.3942 ± 0.0096	82.46 ± 0.79
HB	$\beta = 0.6$	76.19 ± 0.33	99.40 ± 0.15	0.9429 ± 0.0039	0.6052 ± 0.0240	0.3964 ± 0.0120	82.15 ± 0.98
HB	$\beta = 0.65$	76.18 ± 0.24	99.43 ± 0.19	0.9447 ± 0.0033	0.6196 ± 0.0365	0.4008 ± 0.0094	82.26 ± 0.98
HB	$\beta = 0.7$	76.33 ± 0.23	99.42 ± 0.17	0.9449 ± 0.0031	0.6288 ± 0.0363	0.4022 ± 0.0054	82.13 ± 0.72
HB	$\beta = 0.75$	76.35 ± 0.35	99.41 ± 0.12	0.9460 ± 0.0039	0.6208 ± 0.0422	0.4039 ± 0.0089	82.28 ± 0.46
HB	$\beta = 0.8$	76.49 ± 0.35	99.41 ± 0.16	0.9459 ± 0.0032	0.6147 ± 0.0424	0.4025 ± 0.0084	82.43 ± 0.58
HB	$\beta = 0.85$	76.61 ± 0.35	99.45 ± 0.22	0.9462 ± 0.0030	0.6258 ± 0.0418	0.4027 ± 0.0093	82.46 ± 0.70
HB	$\beta = 0.9$	76.58 ± 0.21	**99.47 ± 0.20**	0.9457 ± 0.0025	0.6039 ± 0.0428	0.4003 ± 0.0151	82.53 ± 0.90
HB	$\beta = 0.95$	76.69 ± 0.18	99.44 ± 0.16	**0.9469 ± 0.0029**	0.6073 ± 0.0369	0.3989 ± 0.0097	**82.73 ± 0.60**
SEAL	$I = 1$	74.28 ± 0.42	98.83 ± 0.28	**0.7838 ± 0.0039**	0.4964 ± 0.0322	0.2015 ± 0.0030	85.35 ± 1.07
SEAL	$I = 2$	71.97 ± 0.33	**98.92 ± 0.23**	0.6846 ± 0.0036	0.4379 ± 0.0102	0.1404 ± 0.0018	**85.48 ± 0.65**
SEAL	$I = 3$	69.16 ± 0.61	98.28 ± 0.38	0.5987 ± 0.0050	0.4109 ± 0.0141	0.1055 ± 0.0027	83.82 ± 0.73
SEAL	$I = 4$	66.09 ± 0.23	97.08 ± 0.27	0.5291 ± 0.0032	0.3794 ± 0.0107	0.0866 ± 0.0018	81.33 ± 0.73
SEAL	$I = 5$	63.62 ± 0.64	94.99 ± 1.02	0.4667 ± 0.0018	**0.3379 ± 0.0107**	**0.0761 ± 0.0025**	76.91 ± 0.75
PLC	$\theta_0 = 0.2, \beta = 0.05$	76.51 ± 0.27	**99.33 ± 0.20**	0.9469 ± 0.0033	0.6126 ± 0.0424	**0.4042 ± 0.0119**	**82.37 ± 0.72**
PLC	$\theta_0 = 0.2, \beta = 0.1$	76.15 ± 0.30	99.27 ± 0.17	**0.9484 ± 0.0016**	**0.6073 ± 0.0424**	0.4200 ± 0.0051	81.55 ± 0.55
PLC	$\theta_0 = 0.2, \beta = 0.2$	75.48 ± 0.29	99.22 ± 0.20	0.9466 ± 0.0024	0.6417 ± 0.0374	0.4379 ± 0.0140	80.46 ± 0.91
PLC	$\theta_0 = 0.2, \beta = 0.3$	74.17 ± 0.40	98.93 ± 0.31	0.9454 ± 0.0039	0.6093 ± 0.0287	0.4646 ± 0.0128	78.47 ± 0.79
PLC	$\theta_0 = 0.3, \beta = 0.05$	75.50 ± 0.79	98.91 ± 0.24	0.9448 ± 0.0030	0.6479 ± 0.0258	0.4209 ± 0.0089	80.85 ± 1.05
PLC	$\theta_0 = 0.3, \beta = 0.1$	75.37 ± 0.81	98.93 ± 0.19	0.9459 ± 0.0040	0.6469 ± 0.0476	0.4345 ± 0.0149	80.30 ± 1.50
PLC	$\theta_0 = 0.3, \beta = 0.2$	74.84 ± 1.05	98.93 ± 0.27	0.9458 ± 0.0038	0.6528 ± 0.0439	0.4473 ± 0.0228	79.61 ± 1.87
PLC	$\theta_0 = 0.3, \beta = 0.3$	73.79 ± 1.39	98.84 ± 0.35	0.9450 ± 0.0035	0.6774 ± 0.0488	0.4715 ± 0.0301	77.60 ± 2.43
PLC	$\theta_0 = 0.4, \beta = 0.05$	73.83 ± 0.88	98.27 ± 0.36	0.9412 ± 0.0040	0.6830 ± 0.0150	0.4475 ± 0.0069	79.17 ± 1.51
PLC	$\theta_0 = 0.4, \beta = 0.1$	73.67 ± 0.99	98.26 ± 0.49	0.9415 ± 0.0027	0.6859 ± 0.0254	0.4521 ± 0.0073	79.03 ± 1.32
PLC	$\theta_0 = 0.4, \beta = 0.2$	72.96 ± 0.74	98.25 ± 0.50	0.9441 ± 0.0047	0.6861 ± 0.0484	0.4886 ± 0.0150	77.20 ± 1.00
PLC	$\theta_0 = 0.4, \beta = 0.3$	71.99 ± 0.40	98.17 ± 0.38	0.9443 ± 0.0058	0.6824 ± 0.0321	0.5163 ± 0.0106	75.80 ± 0.83
PLC	$\theta_0 = 0.5, \beta = 0.05$	70.51 ± 1.74	96.70 ± 0.77	0.9383 ± 0.0044	0.7021 ± 0.0203	0.5238 ± 0.0063	73.65 ± 2.62
PLC	$\theta_0 = 0.5, \beta = 0.1$	70.52 ± 1.75	96.67 ± 0.76	0.9385 ± 0.0043	0.7023 ± 0.0245	0.5267 ± 0.0073	73.67 ± 2.61
PLC	$\theta_0 = 0.5, \beta = 0.2$	70.47 ± 1.74	96.63 ± 0.86	0.9361 ± 0.0071	0.6996 ± 0.0267	0.5212 ± 0.0165	73.79 ± 2.52
PLC	$\theta_0 = 0.5, \beta = 0.3$	70.40 ± 1.76	96.66 ± 0.84	0.9357 ± 0.0071	0.7032 ± 0.0302	0.5267 ± 0.0168	73.48 ± 2.50
OLS	$\alpha = 0.1$	67.94 ± 0.39	92.22 ± 0.77	0.3917 ± 0.0036	**0.3214 ± 0.0070**	**0.0610 ± 0.0011**	75.27 ± 1.14
OLS	$\alpha = 0.2$	72.03 ± 0.68	96.23 ± 0.52	0.5635 ± 0.0026	0.4285 ± 0.0099	0.1086 ± 0.0017	80.04 ± 0.93
OLS	$\alpha = 0.3$	73.89 ± 0.38	97.57 ± 0.41	0.6823 ± 0.0052	0.4826 ± 0.0143	0.1638 ± 0.0051	81.15 ± 0.92
OLS	$\alpha = 0.4$	74.82 ± 0.43	98.18 ± 0.31	0.7667 ± 0.0069	0.5151 ± 0.0133	0.2149 ± 0.0100	82.45 ± 0.98
OLS	$\alpha = 0.5$	75.42 ± 0.35	98.64 ± 0.32	0.8274 ± 0.0037	0.5397 ± 0.0167	0.2643 ± 0.0085	82.30 ± 1.05
OLS	$\alpha = 0.6$	75.81 ± 0.28	98.93 ± 0.23	0.8623 ± 0.0090	0.5605 ± 0.0155	0.2955 ± 0.0144	82.47 ± 0.60
OLS	$\alpha = 0.7$	76.17 ± 0.24	99.05 ± 0.25	0.8938 ± 0.0078	0.5730 ± 0.0117	0.3304 ± 0.0096	82.46 ± 0.68
OLS	$\alpha = 0.8$	76.35 ± 0.19	99.16 ± 0.15	0.9181 ± 0.0036	0.5780 ± 0.0258	0.3604 ± 0.0099	82.60 ± 0.58
OLS	$\alpha = 0.9$	76.63 ± 0.14	**99.30 ± 0.17**	**0.9336 ± 0.0041**	0.5852 ± 0.0273	0.3774 ± 0.0101	**82.90 ± 0.57**

Table 3: Performance of soft label methods with different configurations.

References

[1] Asako Kanezaki, Yasuyuki Matsushita, and Yoshifumi Nishida. Rotationnet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 5010–5019, 2018.

[2] Asako Kanezaki, Yasuyuki Matsushita, and Yoshifumi Nishida. Rotationnet for joint object categorization and unsupervised pose estimation from multi-view images. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 43(1):269–283, 2019.

[3] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox. A large-scale hierarchical multi-view rgb-d object dataset. In *IEEE Conf. Robot. Automat. (ICRA)*, pages 1817–1824, 2011.

[4] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 1912–1920, 2015.

[5] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 1588–1597, 2019.

[6] Xinhai Liu, Zhizhong Han, Yu-Shen Liu, and Matthias Zwicker. Fine-grained 3d shape classification with hierarchical part-view attention. *IEEE Trans. Image Process. (TIP)*, 30:1744–1758, 2021.

[7] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks for 3d shape recognition. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 945–953, 2015.

[8] Edward Johns, Stefan Leutenegger, and Andrew J Davison. Pairwise decomposition of image sequences for active multi-view recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 3813–3822, 2016.

[9] Song Bai, Xiang Bai, Zhichao Zhou, Zhaoxiang Zhang, and Longin Jan Latecki. Gift: A real-time and scalable 3d shape search engine. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 5023–5032, 2016.

[10] Song Bai, Xiang Bai, Zhichao Zhou, Zhaoxiang Zhang, Qi Tian, and Longin Jan Latecki. Gift: Towards scalable 3d shape retrieval. *IEEE Trans. Multimedia (TMM)*, 19(6):1257–1271, 2017.

[11] Chu Wang, Marcello Pelillo, and Kaleem Siddiqi. Dominant set clustering and pooling for multi-view 3d object recognition. In *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2017.

[12] Yifan Feng, Zizhao Zhang, Xibin Zhao, Rongrong Ji, and Yue Gao. Gvcnn: Group-view convolutional neural networks for 3d shape recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 264–272, 2018.

[13] Tan Yu, Jingjing Meng, and Junsong Yuan. Multi-view harmonized bilinear network for 3d object recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 186–194, 2018.

[14] Songle Chen, Lintao Zheng, Yan Zhang, Zhixin Sun, and Kai Xu. Veram: View-enhanced recurrent attention model for 3d shape classification. *IEEE Trans. Vis. Comput. Graphics (TVCG)*, 25(12):3244–3257, 2018.

[15] Zhizhong Han, Mingyang Shang, Zhenbao Liu, Chi-Man Vong, Yu-Shen Liu, Matthias Zwicker, Junwei Han, and CL Philip Chen. Seqviews2seqlabels: Learning 3d global features via aggregating sequential views by rnn with attention. *IEEE Trans. Image Process. (TIP)*, 28(2):658–672, 2018.

[16] Jong-Chyi Su, Matheus Gadelha, Rui Wang, and Subhransu Maji. A deeper look at 3d shape classifiers. In *Proc. Eur. Conf. Comput. Vis. Workshops (ECCVW)*, 2018.

[17] Chao Ma, Yulan Guo, Jungang Yang, and Wei An. Learning multi-view representation with lstm for 3-d shape recognition and retrieval. *IEEE Trans. Multimedia (TMM)*, 21(5):1169–1182, 2018.

[18] Zhengyue Huang, Zhehui Zhao, Hengguang Zhou, Xibin Zhao, and Yue Gao. Deepccfv: Camera constraint-free multi-view convolutional neural network for 3d object retrieval. In *Proc. AAAI Conf. Artif. Intell. (AAAI)*, volume 33, pages 8505–8512, 2019.

[19] Jianwen Jiang, Di Bao, Ziqiang Chen, Xibin Zhao, and Yue Gao. Mlvcnn: Multi-loop-view convolutional neural network for 3d shape retrieval. In *Proc. AAAI Conf. Artif. Intell. (AAAI)*, volume 33, pages 8513–8520, 2019.

[20] Zhizhong Han, Honglei Lu, Zhenbao Liu, Chi-Man Vong, Yu-Shen Liu, Matthias Zwicker, Junwei Han, and CL Philip Chen. 3d2seqviews: Aggregating sequential views for 3d global feature learning by cnn with hierarchical attention aggregation. *IEEE Trans. Image Process. (TIP)*, 28(8):3986–3999, 2019.

[21] Zhizhong Han, Xiyang Wang, Chi Man Vong, Yu-Shen Liu, Matthias Zwicker, and CL Chen. 3dviewgraph: Learning global features for 3d shapes from a graph of unordered views with attention. In *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2019.

[22] He-Yu Zhou, An-An Liu, Wei-Zhi Nie, and Jie Nie. Multi-view saliency guided deep neural network for 3-d object retrieval and classification. *IEEE Trans. Multimedia (TMM)*, 22(6):1496–1506, 2019.

[23] Carlos Esteves, Yinshuang Xu, Christine Allen-Blanchette, and Kostas Daniilidis. Equivariant multi-view networks. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 1568–1577, 2019.

[24] Ze Yang and Liwei Wang. Learning relationships for multi-view 3d object recognition. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 7505–7514, 2019.

[25] Xin Wei, Ruixuan Yu, and Jian Sun. View-gcn: View-based graph convolutional network for 3d shape analysis. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pages 1850–1859, 2020.

[26] Jiaxin Chen, Jie Qin, Yuming Shen, Li Liu, Fan Zhu, and Ling Shao. Learning attentive and hierarchical representations for 3d shape recognition. In *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pages 105–122. Springer, 2020.

[27] Kai Sun, Jiangshe Zhang, Junmin Liu, Ruixuan Yu, and Zengjie Song. Drcnn: Dynamic routing convolutional neural network for multi-view 3d object recognition. *IEEE Trans. Image Process. (TIP)*, 30:868–877, 2020.

[28] Jingjia Huang, Wei Yan, Thomas H Li, Shan Liu, and Ge Li. Learning the global descriptor for 3d object recognition based on multiple views decomposition. *IEEE Trans. Multimedia (TMM)*, 2020.

[29] Jinglin Xu, Xiangsen Zhang, Wenbin Li, Xinwang Liu, and Junwei Han. Joint multi-view 2d convolutional neural networks for 3d object classification. In *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, pages 3202–3208, 2020.

[30] Tan Yu, Jingjing Meng, Ming Yang, and Junsong Yuan. 3d object representation learning: A set-to-set matching perspective. *IEEE Trans. Image Process. (TIP)*, 30:2168–2179, 2021.

[31] Jingjia Huang, Wei Yan, Ge Li, Thomas Li, and Shan Liu. Learning disentangled representation for multi-view 3d object recognition. *IEEE Trans. Circuits Syst. Video Technol. (TCSVT)*, 2021.

[32] Weizhi Nie, Yue Zhao, Dan Song, and Yue Gao. Dan: Deep-attention network for 3d shape recognition. *IEEE Trans. Image Process. (TIP)*, 30:4371–4383, 2021.

[33] Yong Xu, Chaoda Zheng, Ruotao Xua, Yuhui Quan, and Haibin Ling. Multi-view 3d shape recognition via correspondence-aware deep learning. *IEEE Trans. Image Process. (TIP)*, 2021.

[34] Xin Wei, Yifei Gong, Fudong Wang, Xing Sun, and Jian Sun. Learning canonical view representation for 3d shape recognition with arbitrary views. In *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pages 407–416, 2021.

[35] Yue Zhao, Weizhi Nie, An-An Liu, Zan Gao, and Yuting Su. Svhan: Sequential view based hierarchical attention network for 3d shape recognition. In *Proc. ACM Int. Conf. Multimedia (MM)*, pages 2130–2138, 2021.

[36] Shuo Chen, Tan Yu, and Ping Li. Mvt: Multi-view vision transformer for 3d object recognition. In *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2021.

[37] Zehua Liu, Yuhe Zhang, Jian Gao, and Shurui Wang. Vfmvac: View-filtering-based multi-view aggregating convolution for 3d shape recognition and retrieval. *Pattern Recognit. (PR)*, page 108774, 2022.

[38] Zizhao Zhang, Haojie Lin, Xibin Zhao, Rongrong Ji, and Yue Gao. Inductive multi-hypergraph learning and its application on view-based 3d object classification. *IEEE Trans. Image Process. (TIP)*, 27(12): 5957–5968, 2018.

[39] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proc. AAAI Conf. Artif. Intell. (AAAI)*, volume 33, pages 3558–3565, 2019.

[40] Junjie Bai, Biao Gong, Yining Zhao, Fuqiang Lei, Chenggang Yan, and Yue Gao. Multi-scale representation learning on hypergraph for 3d shape retrieval and recognition. *IEEE Trans. Image Process. (TIP)*, 2021.

[41] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. Hgnn+: General hypergraph neural networks. *IEEE Trans. Pattern Anal. Mach. Intell. (TPAMI)*, 2022.

[42] Liping Nong, Jie Peng, Wenhui Zhang, Jiming Lin, Hongbing Qiu, and Junyi Wang. Adaptive multi-hypergraph convolutional networks for 3d object classification. *IEEE Trans. Multimedia (TMM)*, 2022.

[43] Jiying Zhang, Fuyang Li, Xi Xiao, Tingyang Xu, Yu Rong, Junzhou Huang, and Yatao Bian. Hypergraph convolutional networks via equivalency between hypergraphs and undirected graphs. In *Int. Conf. Mach. Learn. Workshops (ICMLW)*, 2022.

[44] Zhizhong Han, Xinhai Liu, Yu-Shen Liu, and Matthias Zwicker. Parts4feature: Learning 3d global features from generally semantic parts in multiple views. In *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2019.

[45] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. In *Adv. Neural Inf. Process. Syst. Workshops (NeurIPSW)*, 2014.

[46] Scott Reed, Honglak Lee, Dragomir Anguelov, Christian Szegedy, Dumitru Erhan, and Andrew Rabinovich. Training deep neural networks on noisy labels with bootstrapping. In *Proc. Int. Conf. Learn. Represent. Workshops (ICLRW)*, 2015.

[47] Pengfei Chen, Junjie Ye, Guangyong Chen, Jingwei Zhao, and Pheng-Ann Heng. Beyond class-conditional assumption: A primary attempt to combat instance-dependent label noise. In *Proc. AAAI Conf. Artif. Intell. (AAAI)*, volume 35, pages 11442–11450, 2021.

[48] Yikai Zhang, Songzhu Zheng, Pengxiang Wu, Mayank Goswami, and Chao Chen. Learning with feature-dependent label noise: A progressive approach. In *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2021.

[49] Chang-Bin Zhang, Peng-Tao Jiang, Qibin Hou, Yunchao Wei, Qi Han, Zhen Li, and Ming-Ming Cheng. Delving deep into label smoothing. *IEEE Trans. Image Process. (TIP)*, 30:5984–5996, 2021.