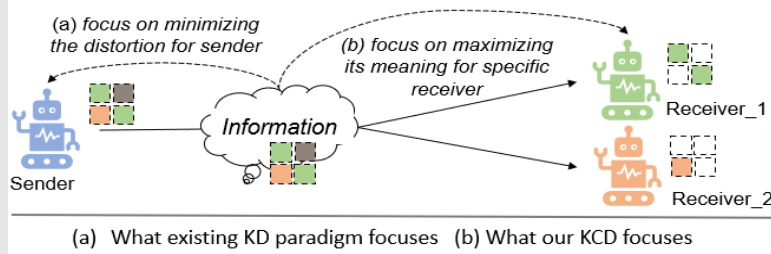
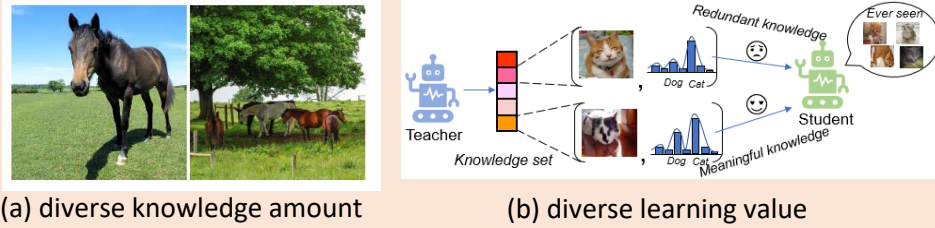


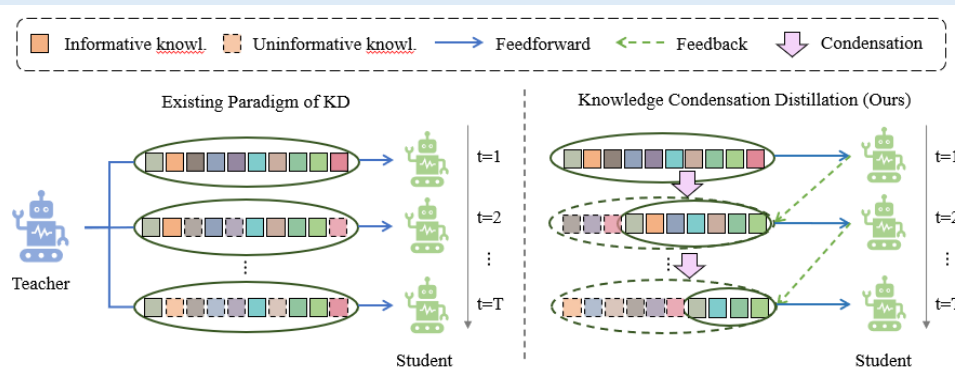
## Background and Motivation

Knowledge shows diverse meaning while is transferred equally



This issue induces the redundancy during KD, which further imposes the challenges:

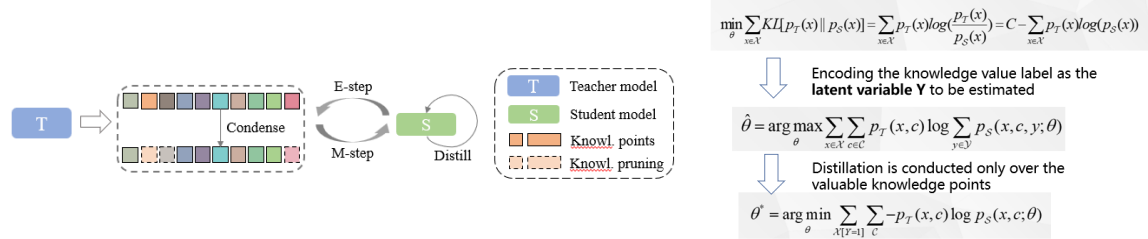
- extra training burden
- worse accuracy performance



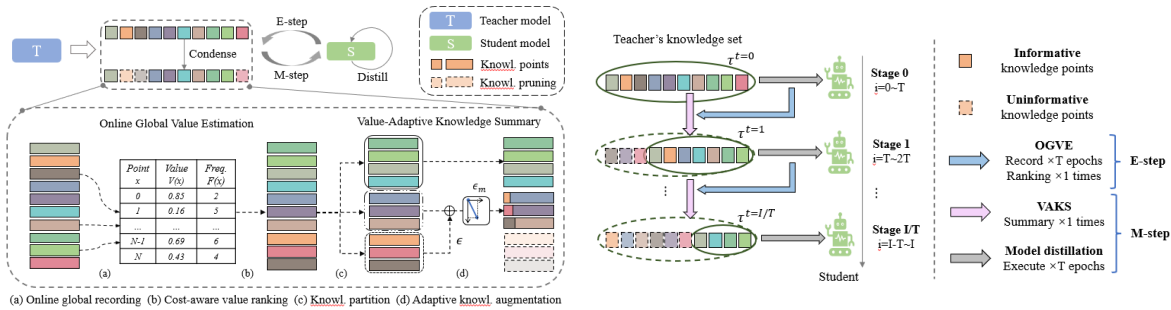
We alleviate this via introducing two novel perspectives:

- Feedback mechanism
- Summary mechanism

## Iterative Optimization based EM formulation



## Overall Framework and Workflow



Teacher	W40-2	W40-2	R56	R32x4	V13	V13	R50	R50	R32x4	R32x4	W40-2
Student	W16-2	W40-1	R20	R8x4	V8	MN2	MN2	V8	SN1	SN2	SN1
Teacher	75.61	75.61	72.34	79.42	74.64	74.64	79.34	79.34	79.42	79.42	75.61
Student	73.26	73.26	69.06	72.50	70.36	64.60	64.60	70.36	70.50	71.82	70.50
KD [10]	74.92	73.54	70.66	73.33	72.98	67.37	67.35	73.81	74.07	74.45	74.83
FitNet [23]	73.58	72.24	69.21	73.50	71.02	64.14	63.16	70.69	73.59	73.54	73.73
AT [13]	74.08	72.77	70.55	73.44	71.43	59.40	58.58	71.84	71.73	72.73	73.32
SP [30]	73.83	72.43	69.67	72.94	72.68	66.30	68.08	73.34	73.48	74.56	74.52
VID [1]	74.11	73.30	70.38	73.09	71.23	65.56	67.57	70.30	73.38	73.40	73.61
RKD [21]	73.35	72.22	69.61	71.90	71.48	64.52	64.43	71.50	72.28	73.21	72.21
PKT [22]	74.54	73.45	70.34	73.64	72.88	67.13	66.52	73.01	74.10	74.69	73.89
CRD [28]	75.64	74.38	71.63	75.46	74.29	69.94	69.54	74.58	75.12	76.05	76.27
WCoRD [9]	76.11	74.72	71.92	76.15	74.72	70.02	70.12	74.68	75.77	76.48	76.68
ReviewKD [4]	76.12	75.09	71.89	75.63	74.84	70.37	69.89	77.45	77.78	77.14	77.14
SSKD [32]	75.66	75.27	70.96	75.80	75.12	70.92	71.14	75.72	77.91	78.37	76.92
KC-KD	75.70	73.84	70.75	74.05	73.44	68.61	67.94	74.41	74.33	75.19	75.60
	(+0.78)	(+0.30)	(+0.09)	(+0.72)	(+0.46)	(+1.24)	(+0.59)	(+0.60)	(+0.26)	(+0.74)	(+0.77)
KC-PKT	75.01	74.12	72.08	74.45	72.82	67.99	67.92	73.32	74.60	75.79	75.78
	(+0.47)	(+0.67)	(+1.74)	(+0.81)	(-0.06)	(+0.86)	(+1.40)	(+0.31)	(+0.50)	(+1.10)	(+1.89)
KC-CRD	75.93	74.60	72.11	75.78	74.38	69.90	69.82	74.49	75.74	76.44	76.4
	(+0.29)	(+0.22)	(+0.48)	(+0.32)	(+0.09)	(-0.04)	(+0.28)	(-0.09)	(+0.62)	(+0.39)	(+0.13)
KC-SSKD	76.24	75.35	71.31	76.48	74.93	71.32	71.29	75.65	78.28	78.59	77.61
	(+0.58)	(+0.08)	(+0.35)	(+0.68)	(-0.21)	(+0.40)	(+0.15)	(-0.07)	(+0.37)	(+0.22)	(+0.69)

## Experiments

Please refer to our ArXiv paper ( <https://arxiv.org/abs/2207.05409> ) as well as the released code ( <https://github.com/dzy3/KCD> ) for more details. Thank you!