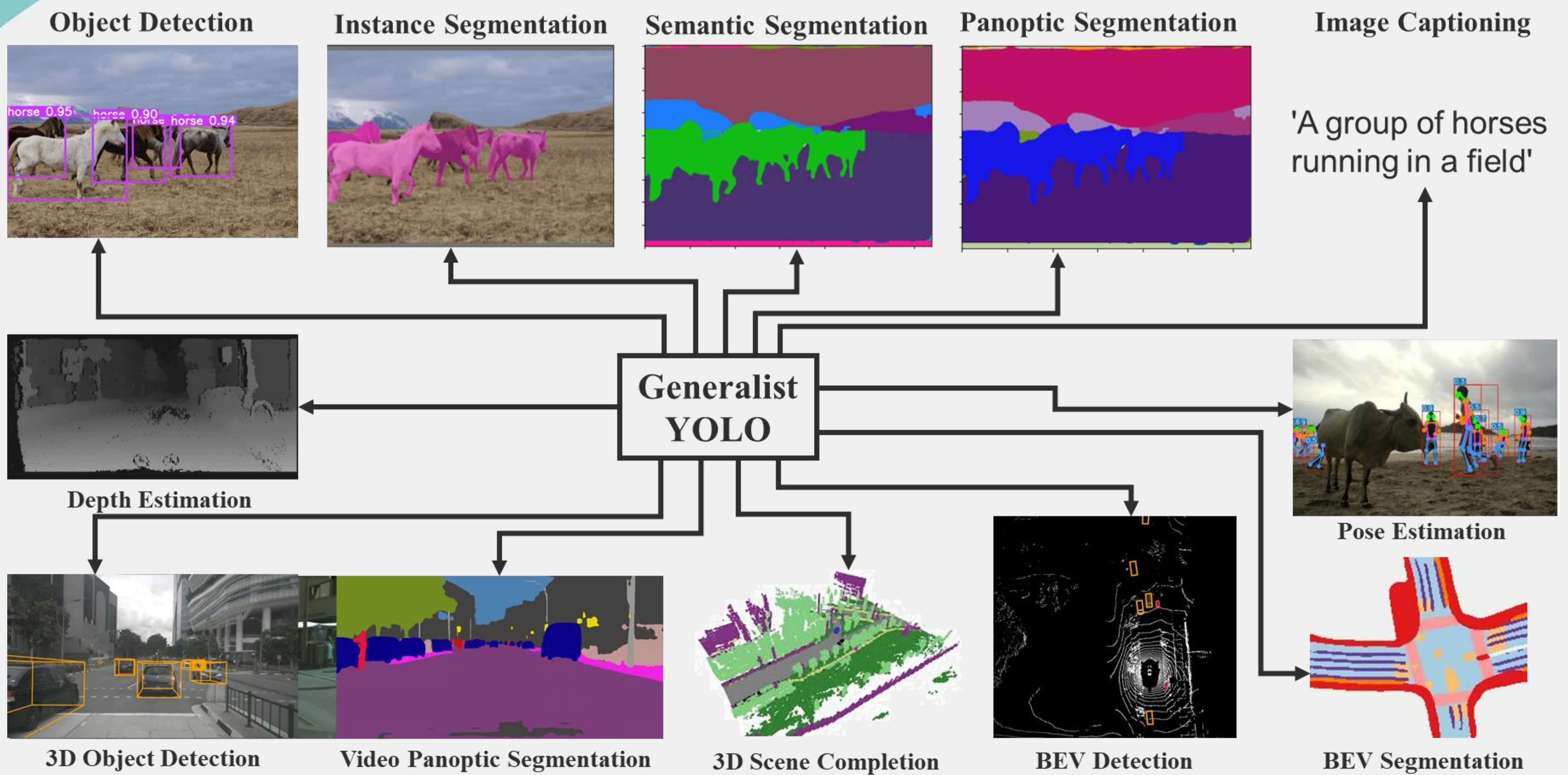


子計畫名稱： 3M Maker: 使用多模態、多任務、
多實例學習開發可信賴多媒體處理技術

子計畫主持人：廖弘源特聘研究員/中央研究院

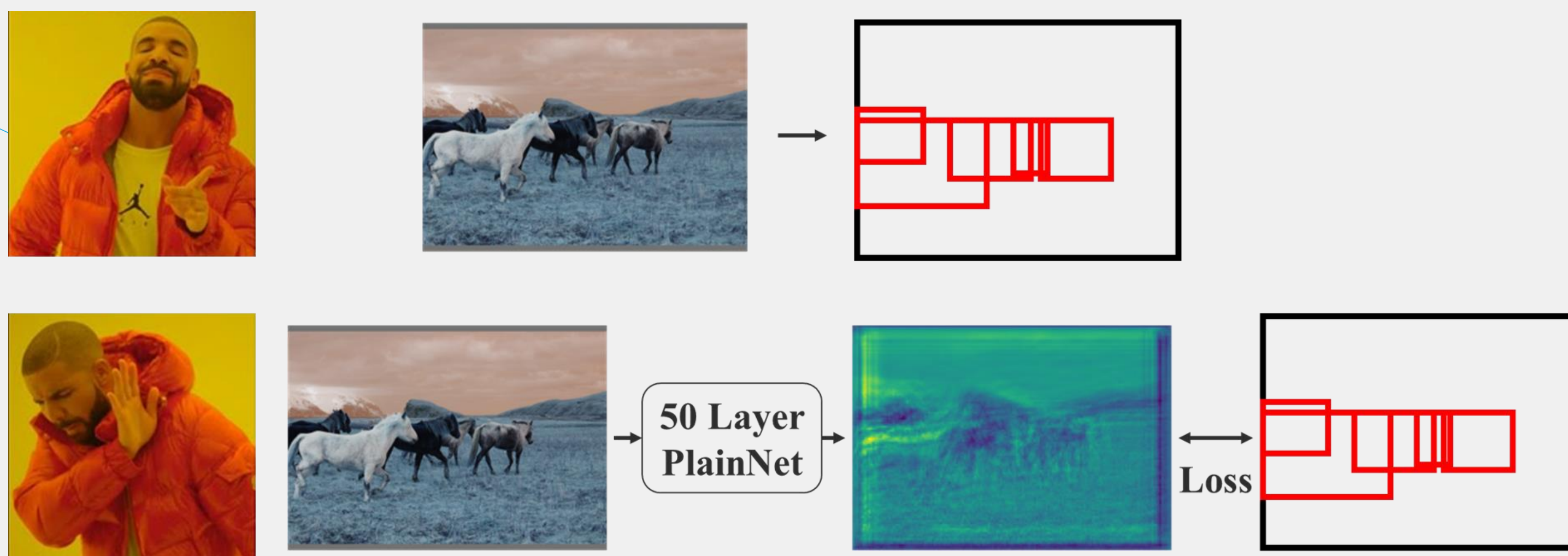
展示技術： Generalist YOLO



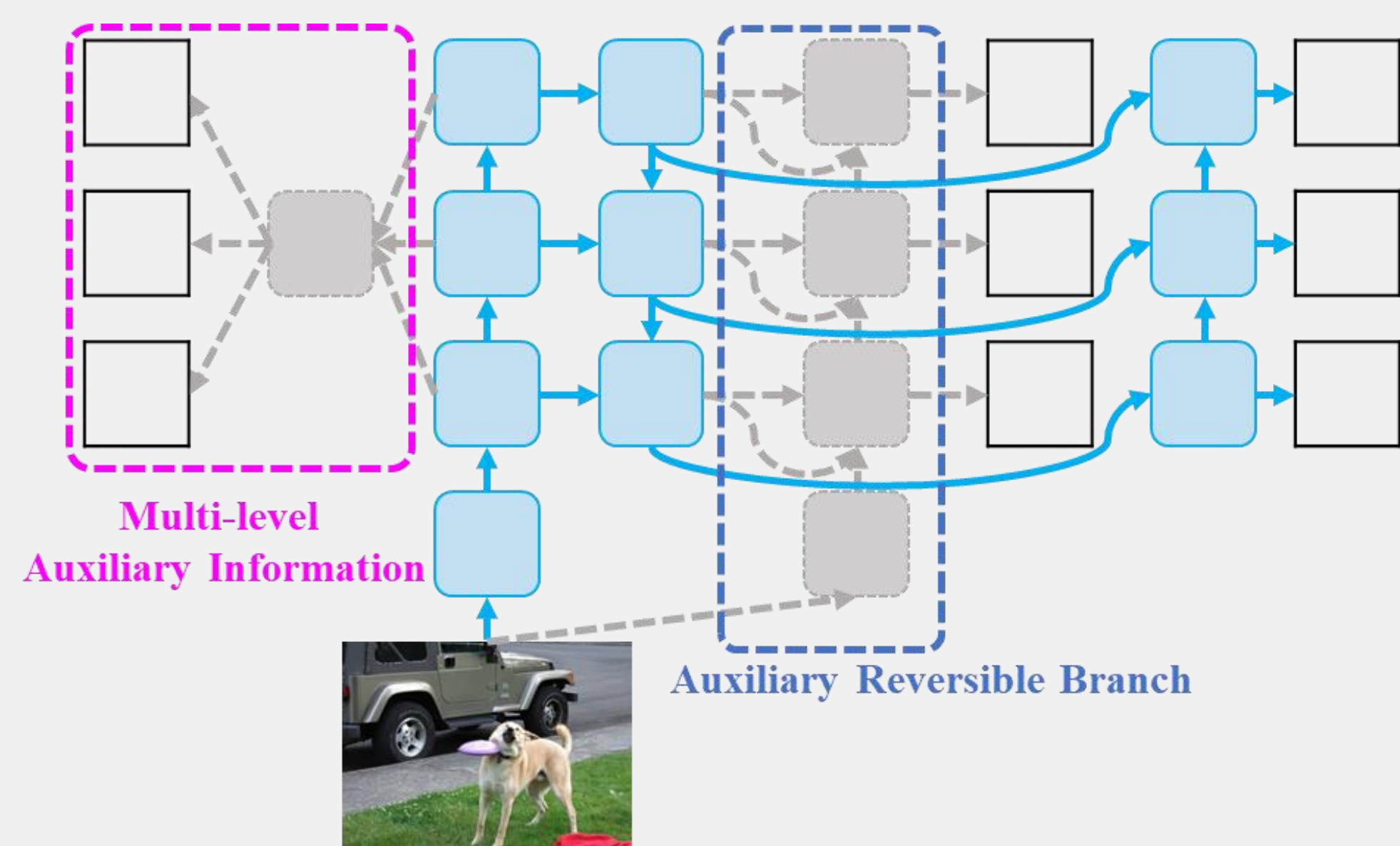
核心技術一 Programmable Gradient Information

資訊瓶頸會導致深度神經網路在前饋階段丟失重要資訊，導致訓練過程以錯誤資訊將輸入與輸出建立不可信賴的關聯。

$$H(X) > I(X, f_1(X)) > I(X, f_2(f_1(X))) > \dots > I(X, F_\theta(X))$$



我們設計了可規劃梯度資訊來確保深度神經網路在前饋階段能保留正確的重要資訊，以此訓練可信賴的深度神經網路。



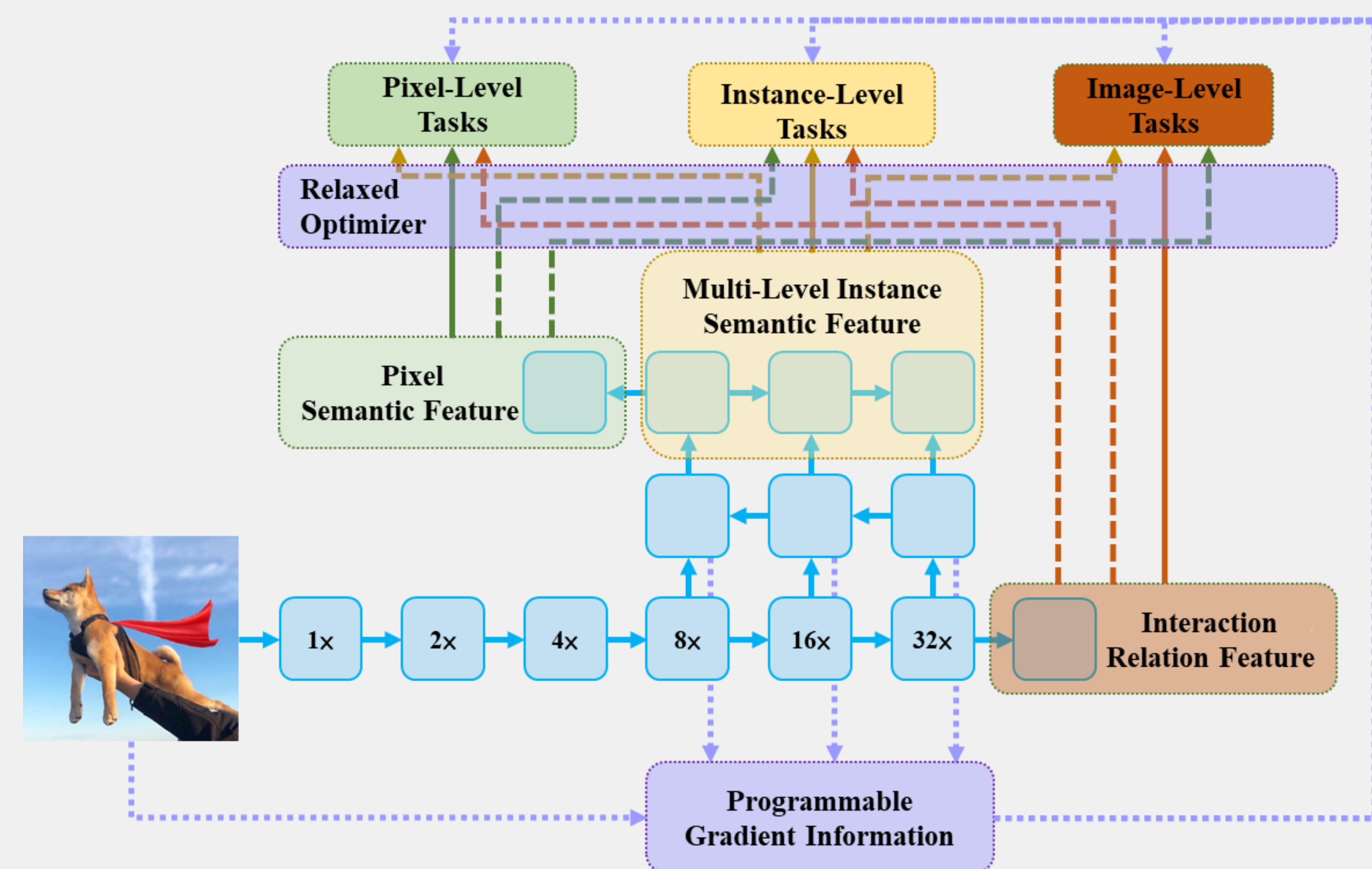
Programmable Gradient Information

可規劃梯度資訊包含兩個元件。(1)輔助可逆分支：將影像重複輸入深層的層並進行監督訓練，藉此找回主架構中深層轉換丟失的重要原始資訊；(2)多層次輔助資訊：確保將所有目標的輔助資訊傳送到每一層，藉此避免中間層的特徵對部分目標產生偏見。

	Detection	Segment	Panoptic	Caption
metric	AP ^{box}	Ap ^{box} / AP ^{seg}	mIoU / PQ	BLEU4
without PGI	52.5	52.3 / 42.4	39.0 / 39.4	38.8
with PGI	53.0	52.9 / 43.2	39.8 / 40.5	39.1

核心技術二 Real-time Multi-Task Vision Generalist

我們將YOLOR擴展至優化器直接對梯度路徑建立先驗知識來驅動多任務學習，並搭配可規劃梯度設計了Generalist YOLO。



Generalist YOLO 比目前最先進方法輕量17倍並快了135倍。

Model	#Param.	AP ^{box}	AP ^{mask}	IoU ^{sem}	IoU ^{stuff}	PQ ^{pan}	B@4 ^{cap}	Acc ^{gnd}
GiT-B [ECCV'24]	131M	46.7	31.9	47.8*	-	-	35.4 ⁺	85.8
GiT-L [ECCV'24]	387M	51.3	35.1	50.6*	-	-	35.7 ⁺	88.4
GiT-H [ECCV'24]	756M	52.9	35.8	52.4*	-	-	36.2 ⁺	89.2
Generalist YOLO	45.5M	52.2	42.9	49.4	56.8	42.2	39.4	-

Model	T ^{total}	T ^{box}	T ^{mask}	T ^{sem}	T ^{stuff}	T ^{pan}	T ^{cap}	T ^{gnd}
GiT-B [ECCV'24]	2589ms	359	1149	717	-	-	272	92
GiT-L [ECCV'24]	5320ms	689	2451	1617	-	-	424	139
GiT-H [ECCV'24]	8321ms	1053	3838	2703	-	-	550	177
Generalist YOLO	61.5ms				61.5			-

看更多:

<https://github.com/WongKinYiu/yolov9>

<https://arxiv.org/abs/2402.13616>