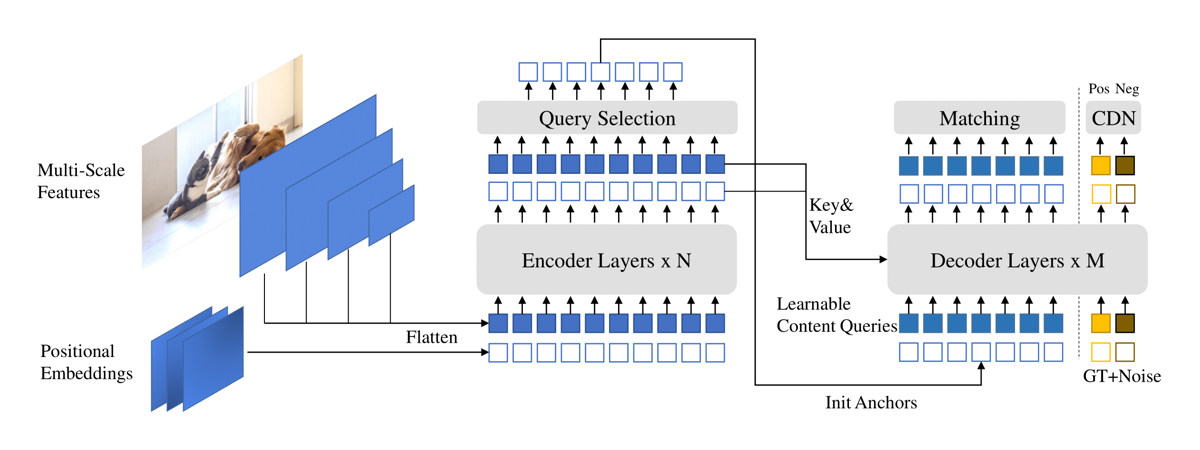
CVPDL-HW1 ( R12944047邱彥慈 )

* Architecture : DINO
  + A DERT-like model composed of backbone + transformer encoders & decoders
    - From previous models
      * Backbone architecture
        + Multiscale features from ResNet, Swin transformer,…
      * Deformable attention
      * Positional query:
        + 4D anchor box dynamically update through each decoder layer
    - Innovation
      * Mixed Query Selection
        + Select top-K encoder features in the last layer to initialize the positional queries, while the content queries are kept learnable as before.
      * Look forward twice
        + In look forward once, the predicted box bi(pred) = f(bi-1, ∆bi), while in look forward twice, ∆bi is used to update the box twice, i.e., bi’ and bi+1(pred).
      * CDN
        + Contrastive Denoising box with both positive and negative samples.



* Implementation
  + Parameter Setting:
    - Start from pretrained weights
      * DINO-4scale provides 3 kinds of pretrained wights, which have been trained on COCO for 12, 24, 36 epochs respectively.   
        I take the 36-epochs one, which is actually checkpoint0029.pth

|  |  |  |  |
| --- | --- | --- | --- |
| learning\_rate | random\_seed | batch\_size | epochs |
| 1e-5 | 0 | 1 | 15 |

* + - Parameters settings are same as the default, except for
      * I’ve trained 24 epochs on ResNet-DINO and found there’s a performance gap at about epoch 12, thus I prefer training to be more than 12 epochs, and it turns out that Swin-DINO takes around 10 epochs to converge.
* Performance
  + Performance of Swin-DINO on validation set

|  |  |  |
| --- | --- | --- |
| AP | AP50 | AP75 |
| 0.586 | 0.859 | 0.614 |
|  | | |

* + Performance on validation set

|  |  |
| --- | --- |
| Swin + DINO | ResNet50 + DINO |
|  |  |

* Visualization
  + On Testing Set
    - Link : <https://github.com/irisowo/CVPDL-HW1/tree/main/DINO/figs/imgs>
    - Preview

|  |  |  |
| --- | --- | --- |
|  | Success Case (Swin) | Failure Case (Res24+24) |
| 5 shark |  |  |

|  |  |  |
| --- | --- | --- |
|  | Failure Case (Swin) | Failure Case (Res24+24) |
| 5 shark |  |  |