

# Medical Image Registration Based on Uncoupled Learning and Accumulative Enhancement

## 1、Problem

医学图像配准是医学图像分析领域中重要的一部分，绝大多数的医学图像分析都需要进行配准的预处理，将研究图像对齐到标准空间以便于数据的批量处理。传统的配准方法自由度较大，需要的超参数较多，结果不够鲁棒。目前基于深度学习的方法虽然能够通过推理过程得到较为不错的配准效果，但是针对于大变形的图像对的效果就会变成较差。因此作者针对于大变形配准提出了一种非解耦学习和累计增强的网络模型。

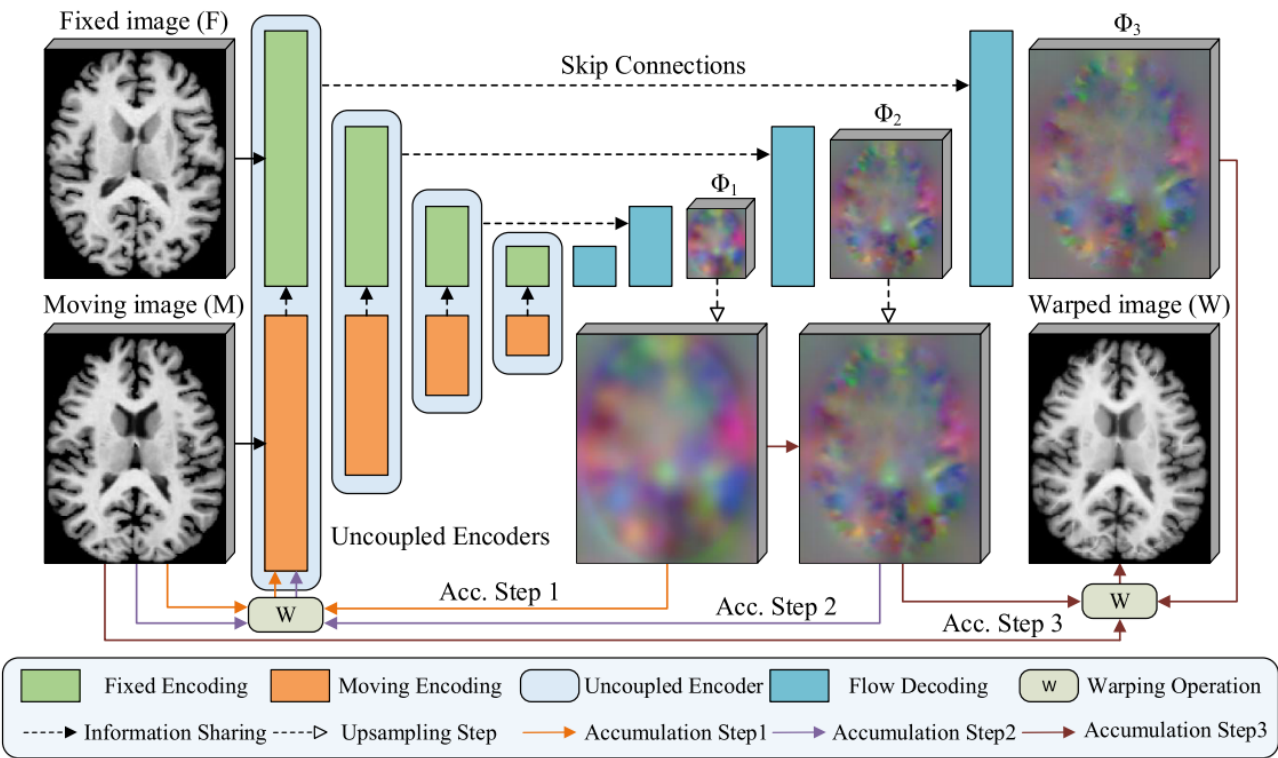
## 2、Contribution

作者针对于大变形配准做了如下的贡献：

- 作者提出了一种非解耦学习和累计增强的网络模型。
- 作者采用了ASPP网络模块对Moving Image做了更加全局的特征提取
- 作者提出的方法相对于baseline有着显著的提升。

## 3、Method

作者提出了一种非解耦学习和累计增强的网络模型如下所示：

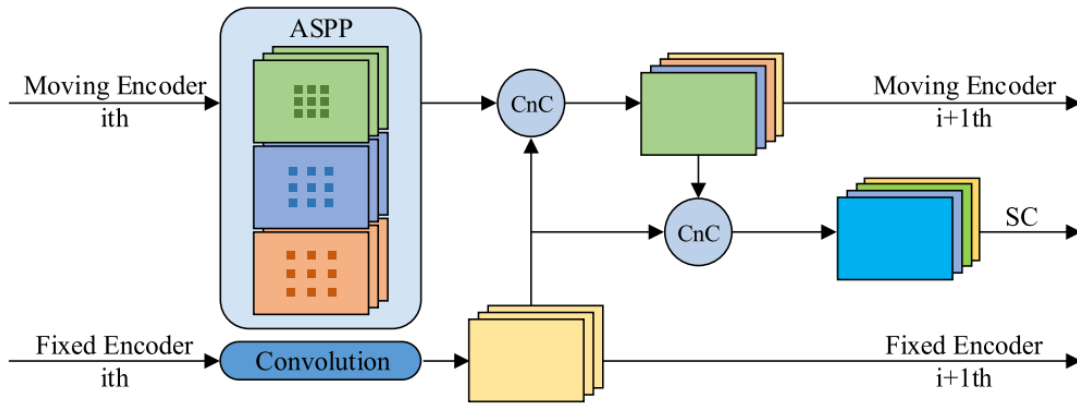


相较于baseline（例如VoxelMorph）而言，网络的输入虽然也采用了Image pairs的方式，但却不是简单的将Fixed Image( $I_F$ )和Moving Image( $I_M$ )进行简单的拼接成双通道然后送入Encoder，而是采用了Uncoupled Encoding的方式去有选择性地学习Image pairs的信息，并且通过Information Sharing的方式将 $I_M$ 的信息与 $I_F$ 的信息共享。

除此以外，还采用了类似Unet的网络框架，对Encoder进行解码，通过设计Uncouple Encoder中的信息共享输出并将这个输出skip connect到Decoder，保留了更多的图像信息，使得之后生成的变形场更加的精细。

作者在此网络架构中提出了累积增强的网络模块，解决了其他方法中粗配准和精配准由于采用了相互独立的Encoder和Decoder而带来的过多的参数学习，并且通过累计增强网络模块能够实现高效的多尺度的粗配准和精配准，充分的利用了Encoder的信息进行变形场的解码。

### 3.1 Uncoupled Spatial Encoder



**Fig. 2.** The Uncoupled Spatial Encoder. CnC denotes Concatenation and Convolution, SC corresponds to Skip Connection.

以上是Uncouple Spatial Encoder的网络结构，可以看出 $I_F$ 和 $I_M$ 并不是简单的图像对的拼接，送入的也不是相同结构的网络，这就使得网络能够更加精细的针对于 $I_M$ 学习变形场，这样有利于处理大变形时 $I_M$ 能够发现和学习更多的远程信息。这里也ASPP模块，通过不同感受野去结合局部和全局的信息，在多尺度上更好的捕获上下文信息。

并且Uncouple Spatial Encoder有着三个不同的输出，其中 $I_F$ 的输出只与其本身有关系，这也保证了配准的 $I_F$ 的信息的不会混入 $I_M$ 的信息，从而使得网络不会学到一些奇怪的变形。

### 3.2 Accumulative Warping Enhancement

作者提出的了累计增强模型也类似于coarse-to-fine的策略，但是这种方法需要的参数量会更加的小，并且也是一种多尺度的变形场的融合。该网络的思想是通过不同的尺度的Encoder的输出（即对变形场的推理）对 $I_M$ 进行扭曲操作分别得到Wrapped Image  $I_{W_1}, I_{W_2}, I_{W_3}$ ，并将三者与 $I_F$ 进行相似度计算。

其中Step1和Step2的变形场 $\phi_1, \phi_2$ 都可以直接由Decoder后的上采样得到，但Step3的变形场 $\phi$ 则是由 $\phi_1$ 和 $\phi_2$ 以及最后一层Decoder的输出 $\phi_3$ ：

$$\phi = S^{2,3}(S^{1,2}(\phi_1) \circ \phi_2 + \phi_2) \circ \phi_3 + \phi_3$$

### 3.3 Loss function

作者提出了一种多Loss function的方式去结合多变形场的学习。 $I_W$ 与 $I_F$ 的相似度计算方式为：

$$L_{sim}(F, M \circ \phi) = - \sum_{i=1}^K \gamma^{i-1} NCC_{w_i}(F, M \circ \phi)$$

最终的多Loss function为：

$$L_{total}(F, M \circ \phi) = L_{sim}(F, M \circ \phi) + \lambda \sum_{p \in \Omega} \|\nabla \phi\|^2$$

因此从Loss function可以看出，通过优化每一个形变场实现图片对的配准。

## 4 Conclusion

作者提出了这种非解耦的方式提取 $I_M$ 的特征，将有利于增强大变形配准的泛化能力，并且模型中没有破坏 $I_F$ 的原始数据的特征提取。使用固定的编码器对特征进行学习，且使用了三个不同尺度的同一迭代方式进行学习是一种很新颖的方式，使得网络变得紧凑和轻量化。

但这篇文章也有一些缺点，比如multi-loss的加入似乎没有提高明显的精度，这里我认为作者应该去做更多的实验证明这种multi-loss的有效性。因为从梯度回传的角度，第一个encoder在一次迭代中会被多次的涉及到。另一个疑惑就是由第一层的Encoder得到 $\phi_1$ 是通过8倍的上采样得到的，这8倍的上采样是否会丢失太多的相关信息？