从近年来图像研究领域的论文主题来看，现在有几大研究方向值得关注。

首先，在算法领域， 生成对抗网络（GAN）的研究呈现井喷的趋势。越来越多的研究者尝试使用 GAN 进行图像方面的研究，如图像识别、 对抗样本攻击和防御，以及生成高清晰度图像、图像风格转换、直接生成新图像等方面的研究。也有部分学者尝试用 GAN 进行 目标检测。

其次，在任务领域，有更多由静态图像转变为动态的视频图像、由 2D 图像研究转向 3D 图像和三维数据方面的研究。近年来，有更多的尝试研究视频领域中的 图像分类、目标分割和检测方面的算法出现，实现了诸如行人检测、人体姿态追踪等方面的应用。同时研究人员尝试使用 深度学习模型探究 3D 建模方面的表现。

最后，在模型方面，出现了模型 参数缩减和模型压缩方面的多种研究。很多学者研究在不影响模型性能的前提下进行模型 剪枝和压缩方面的技术，希望能够将性能优异模型部署在移动端或物联网设备上，实现本地化的模型推断。

“经过R-CNN和Fast RCNN的积淀，Ross B. Girshick在2016年提出了新的Faster RCNN，在结构上，Faster RCNN已经将特征抽取(feature extraction)，proposal提取，bounding box regression(rect refine)，classification都整合在了一个网络中，使得综合性能有较大提高，在检测速度方面尤为明显。”知乎的一段话，说明的是端到端的网络，把模型的各个部分都整合到一个网络中，能具有更好的效果，也就是模型的不同部分一起学习效果会更好。

神经网络中可以添加的子模块和学习目标是可以多种多样的，并且是可以多个子模块一起进行学习的，比如目标检测或者实例分割里面，可以添加检测框学习预测的子模块，它可以学习预测框的大小、位置这些数据。然后和预测分类一起进行学习。

网络训练并不是一定要一次完成的，可以分层不同的的模块进行训练，再合并起来。

目标检测图像检测一类算法的一个启示，最开始即使使用深度学习的目标检测，对于边框都是暴力遍历整个图像所有可能的边框然后再检测目标的。但是仔细想一想既然深度学习具有很强的预测能力，为什么不能用来预测框的位置呢？不过完全直接预测到太难，于是采用anchor的机制，预测小范围内的偏移，这样就是可行的了。

这启示我们在其它任务的过程中也可以将网络的子模块设计成一个预测模块来完成任务，而不是硬编码的模块。

卷积也可以用来预测，不一定非得全连接。比如目标检测，预测偏移量，框的正负得分，每一条特征点-（所有通道的点组成）都预测一个或多个值出来。