# 神经uggets：年龄和性别估计

今天，我们开始了一系列新的帖子，我们称之为neuronugget。2月15日，我们准时发布了第一个版本的。到目前为止，它仍然处于alpha阶段，我们需要花很多时间来实现我们计划的一切。但即使是现在，你也可以做一些很酷的事情。在NeuroNuggets系列中，我们将逐一介绍这些很酷的东西，不仅解释了如何在平台上运行一些东西的技术细节，还解释了每个模型背后的主要思想。这也是我介绍我在俄罗斯圣彼得堡新办公室聘用的新的深度学习团队的机会。

在这篇文章中，我们将介绍我们的第一部分：年龄和性别估计模型。在我们的演示中，这是最简单的神经架构，但即使是这个网络也有很多技巧要解释。我很高兴向大家介绍劳夫·库尔巴诺夫，他是我们在圣彼得堡的第一批雇员之一，我们与他共同撰写了这篇文章：

### 谁雇了个书呆子？

人工智能研究者倾向于质疑直觉的本质。当你问一台计算机如何做对人类来说似乎太简单的事情时，你就会发现，对我们来说“直观清晰”的东西可能很难正式化。我们对人类年龄和性别的视觉感知就是这种微妙品质的一个很好的例子。

对于我们这些人工智能书呆子来说，Elizer Yudkowsky作为人工智能安全研究人员和最受欢迎的哈利波特迷（Harry Potter fanfic）的作者都很熟悉（我们衷心推荐给大家，简称为“HPMoR”）。而《哈利波特》系列为这篇文章提供了一个完美的例子，这是一个虚构的艺术品，看起来直观清晰，但在实践中很难再现：

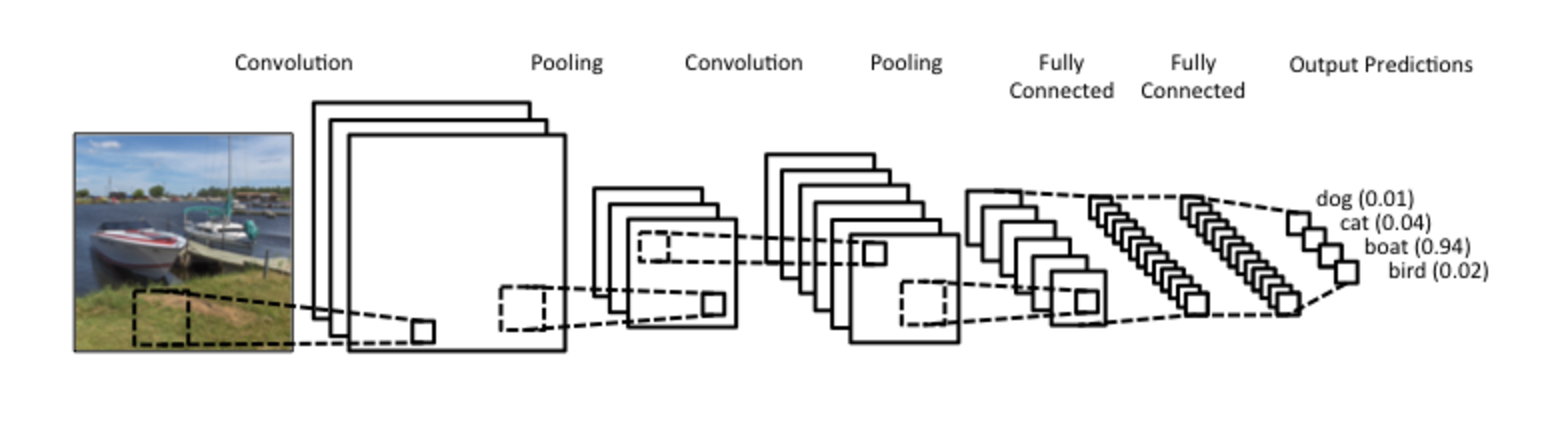


阿不思在附近放置了一个围板，以防止17岁以下的人接近它。年龄线魔法是如此先进，即使是一种老化的药水也无法愚弄它。即使是尤德科夫斯基在HPMoR中并没有以他一丝不苟的方式深入研究年龄线的机制，但今天我们将对此进行一次尝试；在我们讨论这个问题的同时，我们也将尝试性别识别。在计算机视觉中，我们通常从卷积神经网络开始。

### 卷积神经网络

顾名思义，神经网络是一种机器学习方法，它以非常抽象的方式模仿大脑处理信息的方式。它是一个被称为人工神经元或感知器的学习单元网络。在训练过程中，神经元学习如何将输入信号（如猫的图片）转换成相应的输出信号（如本例中的“猫”标签），训练从现实生活中的例子进行自动识别。

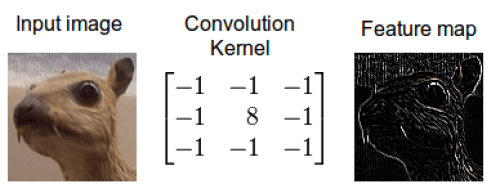
现在几乎所有的计算机视觉都是基于卷积神经网络的。粗略地说，cnn是多层（深层）神经网络，每层在小窗口中处理图像，提取局部特征。渐渐地，一层一层地，局部特征变得全局化，能够从越来越大的原始图像中提取它们的输入。以下是它在一个非常简单的CNN中的工作原理（图片取自，我们建议阅读全文）：



最后，经过几层（有时几百层）之后，我们得到了“看”整个原始图像的全局特征，现在它们可以以相对简单的方式组合以获得类标签（识别是狗、猫、船还是哈利波特）。

从技术上讲，卷积神经网络是一种具有卷积层的神经网络，卷积层是一种将特定的核（滤波器）应用于输入中的每个点（在每个像素中具有多个通道的“图像”，即三维张量）并通过滑动核生成滤波输出的变换输入。

让我们考虑一个过滤器的简单例子：图像中的边缘检测。在这种情况下，用于边缘检测的输入是图像，并且图像中的每个像素由三个数字定义：该像素颜色中的红色、绿色和蓝色的强度。我们构造了一个特殊的内核，它将应用于图像中的每个像素；输出是一个新的“图像”，显示了这个内核的结果。基本上，这里的核心是一个小矩阵。其工作原理如下：

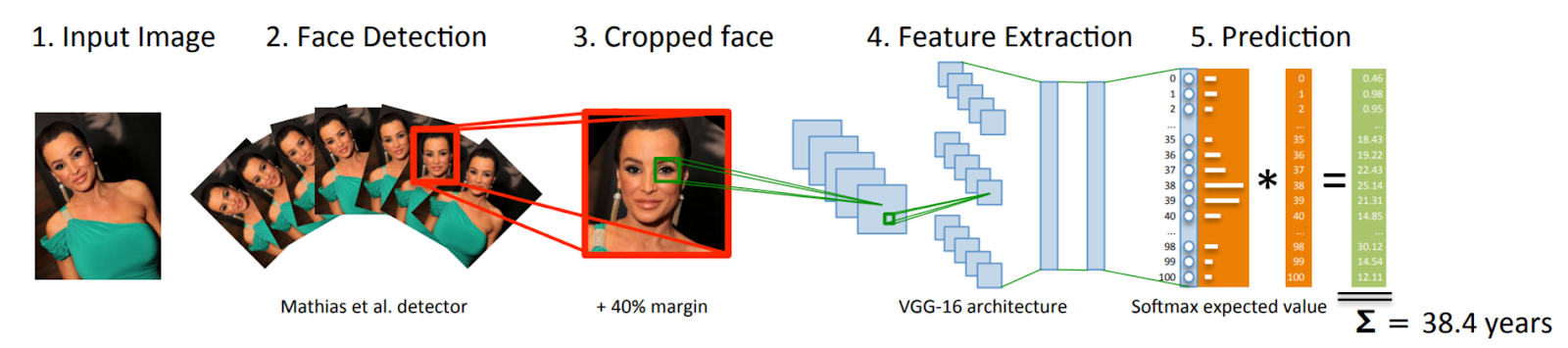


内核在图像中的每个像素上滑动，当有一条边（颜色的突然变化）时，输出值就会增加。在上图中，将这个简单的矩阵元素分别乘以图像中的每个3x3窗口，我们得到了非常好的边缘检测结果。

一旦你了解了滤波器和核函数，解释神经网络中的卷积层就变得非常简单。你可以把它们看作是普通的卷积，就像上面的边缘检测例子一样，但是现在我们在训练网络时端到端地学习卷积核。也就是说，我们不再需要手工发明这些小矩阵，而是可以自动学习矩阵，为特定任务提取最佳特征。

### 模型管道

年龄和性别估计听起来像是一个传统的机器学习任务：对性别进行二元分类（这可能会引发一些争议，但是的，我们的模型生活在一个二元世界）和年龄回归。但在我们开始解决这些问题之前，我们需要找到照片上的面孔！分类不会对整个图片起作用，因为它可能包含多个面。因此，年龄和性别估计问题通常分为两个步骤：人脸检测和被检测人脸的年龄/性别估计：



在你可以在神经平台上找到的模型中，这些步骤是独立执行的，并不是端到端的训练，所以让我们特别讨论每一个步骤。

### 人脸检测

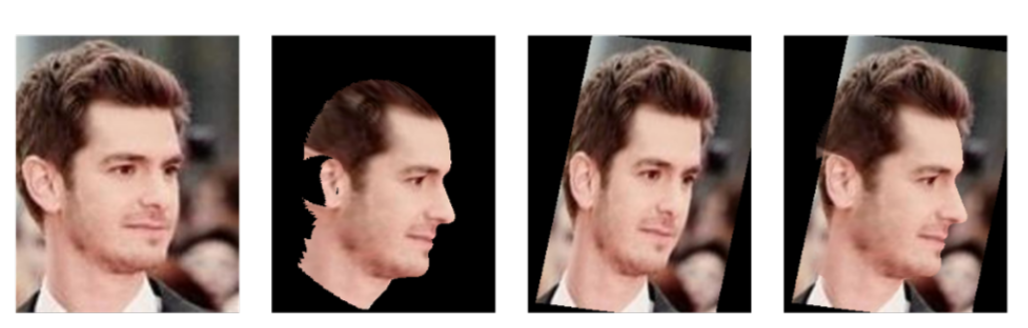
人脸检测是计算机视觉中的一个经典问题。甚至在21世纪初的“深度学习革命”之前，所谓的“深度学习革命”就已经相当成功地解决了这个问题。它是最著名的应用之一，哈尔叶栅作为特点，但那些日子已经过去了…

今天，人脸检测并不是一项单独的任务，需要单独的方法。并用卷积神经网络进行求解。老实说，自从深度学习的出现以来，CNNs在目标检测方面的作用已经很明显了，因此，我们期望一个老问题的现代解决方案也基于CNNs。我们没有错。

但是在现实世界的机器学习中，除了检测精度，你还应该考虑其他的特性，比如简单性和推理速度。如果一个简单的方法足够有效，那么引入非常复杂的模型来获得几个百分点可能是不值得的（提醒我稍后告诉你Netflix大奖挑战赛的结果）。因此，在神经平台演示中，我们使用一种更经典的方法来进行人脸检测，同时保留CNNs作为核心年龄/性别识别任务。但我们也可以从经典的计算机视觉中学到很多。

总之，人脸检测模型可以用HOG+SIFT特征表示的支持向量机来描述。HOG和SIFT表示是手工制作的特征，是多年构建图像识别系统经验的结果。这些特征识别图像局部区域的梯度方向，并执行一系列确定的图像变换。结果表明，这种表示方法在支持向量机（SVM）等核方法中非常有效。

### 数据扩充



在神经信息学，我们是使用计算机视觉合成数据的忠实粉丝。通常，这意味着我们从三维计算机图形学中生成复杂的合成数据集，甚至期望将来使用生成性对抗网络来生成合成数据。但我们不要忘记增加数据集的最基本工具：数据扩充。

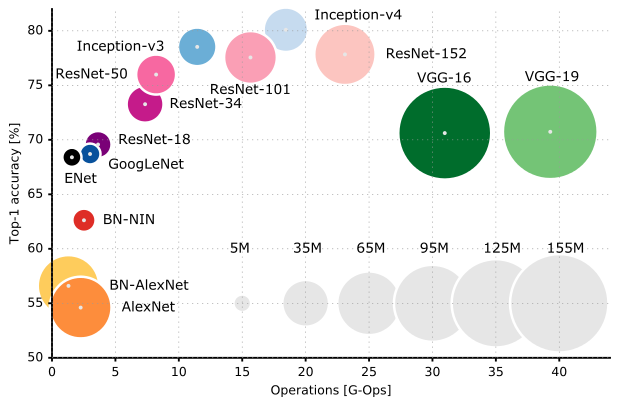
因为我们已经在前面的步骤中提取了人脸，所以只增加人脸就足够了，而不是整个图像。在演示中，我们使用标准的增强技巧，如水平/垂直移动和镜像，同时使用更复杂的随机擦除图像的补丁。

### 年龄估计

为了预测年龄，我们将一个深度卷积神经网络应用于在前一处理阶段检测到的人脸图像。演示中的方法使用（WRN）体系结构，在主流对象检测数据集上优于Inception体系结构，在同一任务上实现两倍更快的收敛。在我们解释什么是剩余网络之前，我们先从一个简短的历史参考开始。

### ImageNet挑战赛

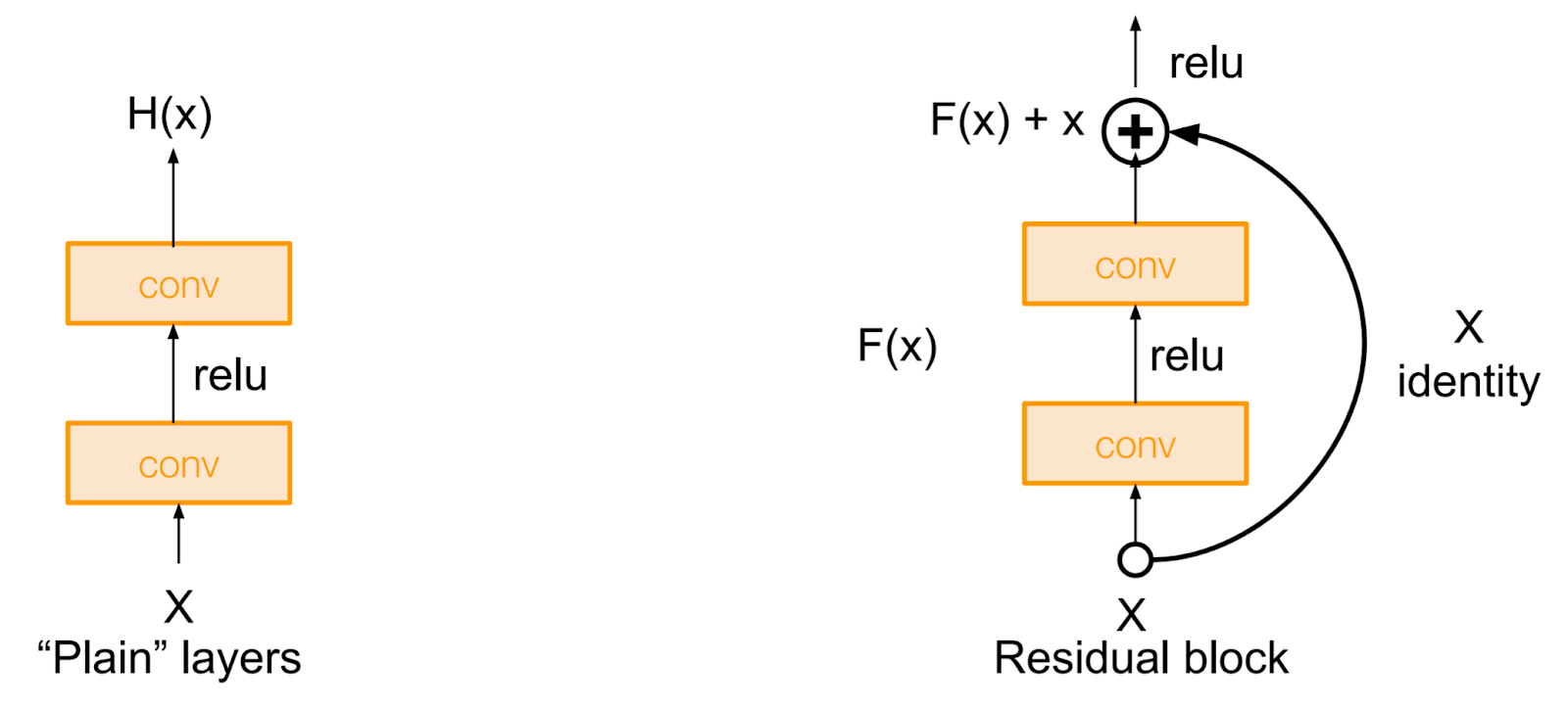
该项目是一个大型视觉设计用于研究。20世纪10年代的革命始于计算机视觉，在解决ImageNet挑战方面取得了巨大进展。ImageNet上的结果不仅在人工智能社区中得到认可，而且在整个行业中也得到认可，ImageNet已经成为并且仍然是最受欢迎的通用计算机视觉数据集。在不涉及太多细节的情况下，让我们先来看看几位第一年获奖者的对比图：



在图上，水平轴显示了模型是如何计算密集的，圆圈大小指示了参数的数量，而垂直轴显示了ImageNet上的图像分类精度。如您所见，ResNet架构显示了一些最好的结果，同时也保持了较好的效率。他们的秘方是什么？

### 残余连接

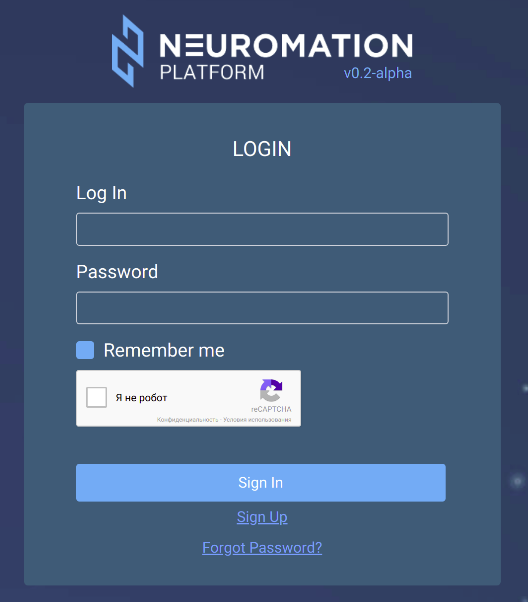
众所周知，较深的模型比较浅的模型表现得更好，它们更具表现力。但是深层模型的优化是一个大问题：深层模型由于梯度从顶层传播到底层的特殊性而难以优化（我希望有一天我们会详细解释）。剩余连接是解决这个问题的一个很好的方法：它们在层的“周围”添加连接，梯度流现在能够在反向传播期间“跳过”过多的层，从而导致更快的收敛和更好的训练：



本质上，与宽残差网络的唯一区别在于，在最初的论文中，我们更仔细地研究了架构的宽度和深度之间的权衡，从而得到了更高效、收敛速度更快的架构。

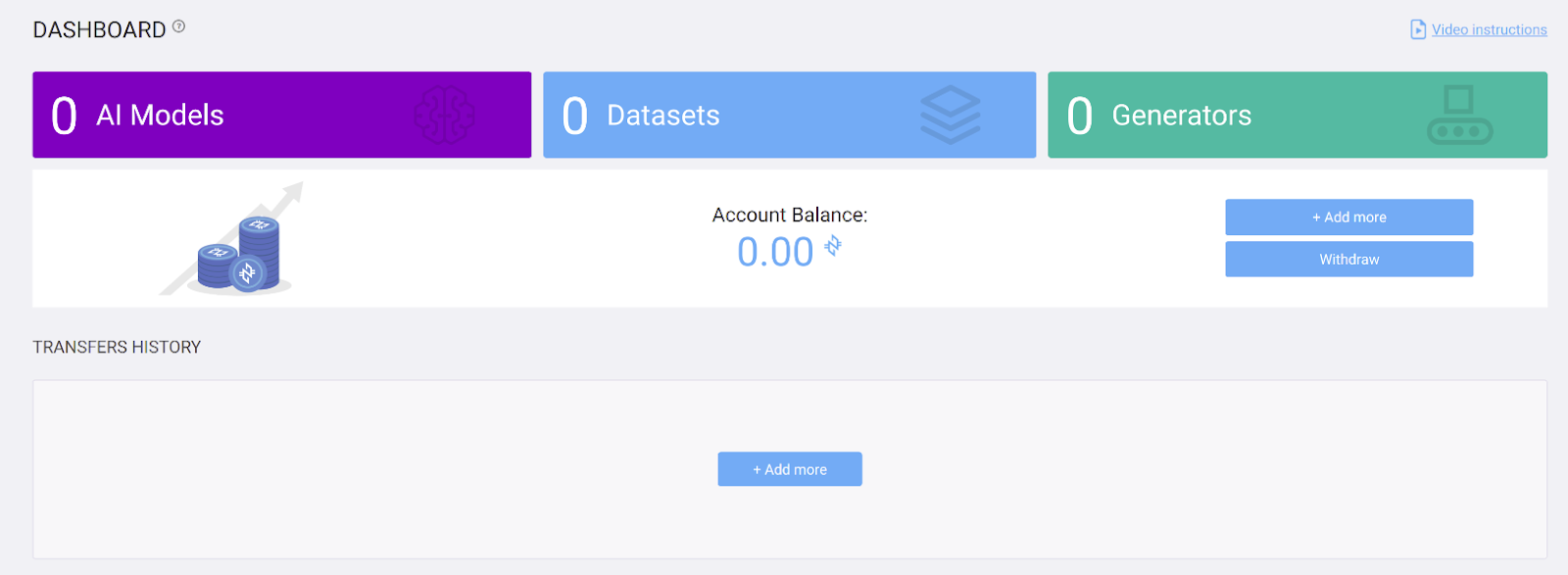
### 神经平台上的DEX

我们将DEX模型包装到docker容器中并上传到神经平台。我们开始吧！首先，进入神经信息登录页面

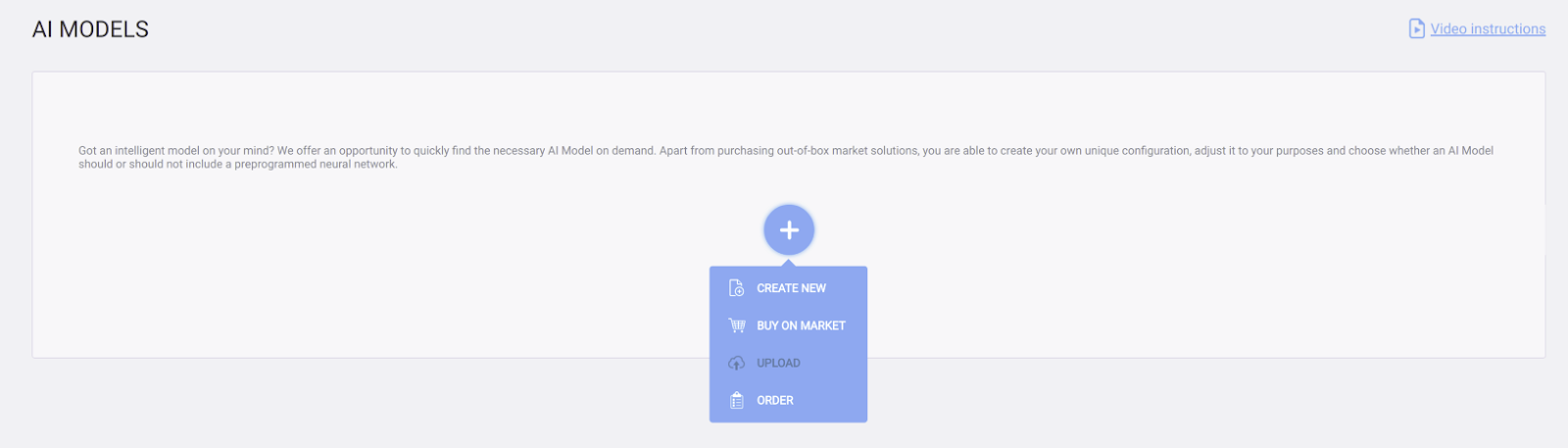


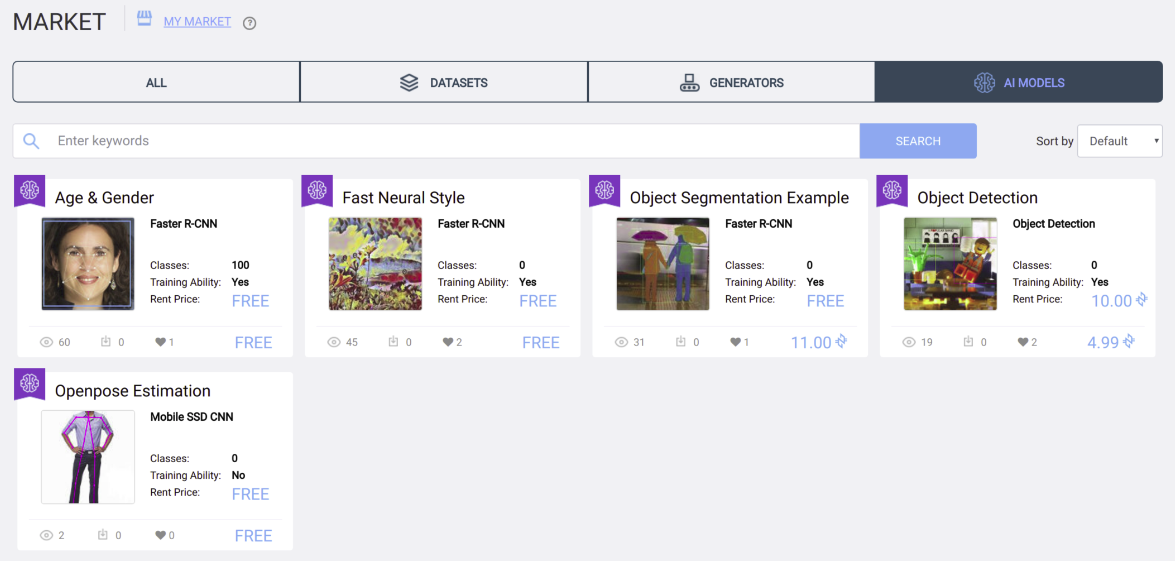
在神经平台上的仪表板上，您可以看到您的余额还剩多少NTK，并将其分为三部分：

* 人工智能模型，
* 数据集，
* 发电机。

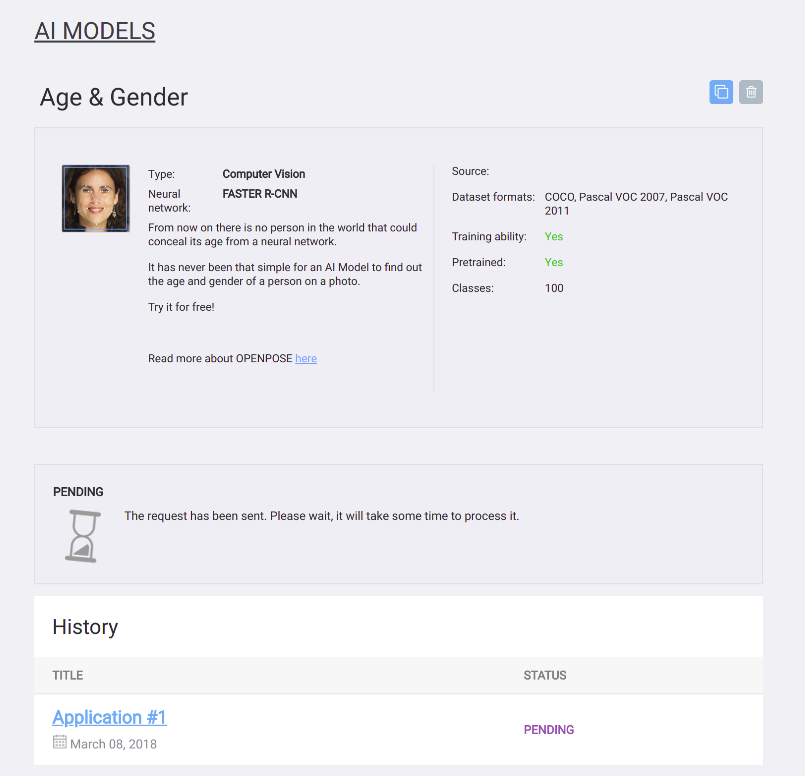


今天我们讨论的是年龄和性别估计器，一个人工智能模型，可在神经市场。让我们购买我们的第一个模型！我们输入人工智能模型并购买年龄和性别模型：

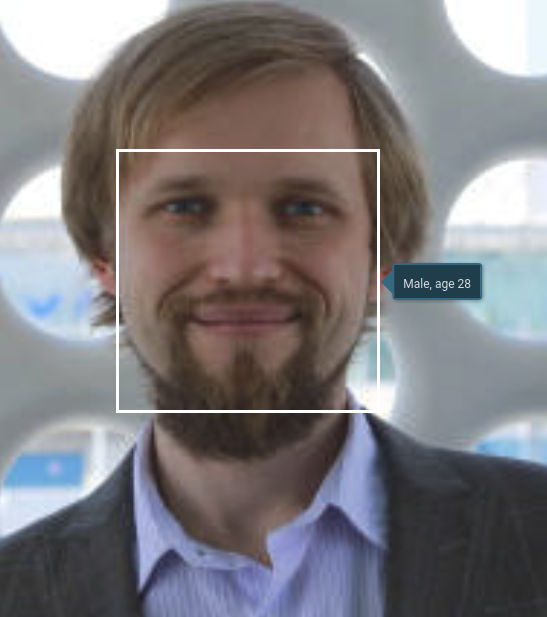




Then we request a new instance; it may take a while:



And here we go! We can now try the model on a demo interface:



It does tend to flatter me a bit.

Sergey NikolenkoChief Research Officer, Neuromation

Rauf KurbanovSenior Researcher, Neuromation