

Garbage Classification

With

Convolutional Neural Networks

Team Members :

Yashraj Bharambe
Diptesh Chaudhari
Pranav Karanjawane
Shrikrushnakumar Sondge
Ashutosh Mishra

Abstract

A computer vision approach to classifying garbage into recycling categories could be an efficient way to process waste. Recycling is already a significant work for all countries. Among the work needed for recycling, garbage classification is the most fundamental step to enable cost efficient recycling.

The objective of this project is to take images of a single piece of recycling or garbage and classify it into six classes consisting of glass, paper, metal, plastic, cardboard, and trash. We attempt to identify a single garbage object in images and classify it into one of the recycling categories. We study several approaches and provide comprehensive evaluation. The models we used include support vector machines (SVM) with HOG features, simple convolutional neural network (CNN), and CNN with residual blocks.

Keywords— Garbage Classification, Deep Learning, Machine Learning, CNN

Table of Contents

Sr No.	
1.	Abstract
2.	Introduction
	a. Problem Statement
3.	Dataset And Data Collection
4.	Model And Methodologies
5.	Approach
6.	Drawbacks

Introduction

Currently, the world generates more than 2 billion tons of municipal solid waste annually, which is huge damage to the ecological environment. Waste production will increase by 70% if current conditions persist. Recycling is becoming an indispensable part of a sustainable society. However, the whole procedure of recycling demands a huge hidden cost, which is caused by selection, classification, and processing of the recycled materials. Even though consumers are willing to do their own garbage sorting nowadays in many countries, they might be confused about how to determine the correct category of the garbage when disposing of a large variety of materials.

Finding an automatic way to do the recycling is now of great value to an industrial and information-based society, which has not only environmental effects but also beneficial economic effects. Since last decade the industry of artificial intelligence has welcomed its third wave with a sufficient database. Deep learning began to show its high efficiency and low complexity in the field of computer vision. Many new ideas were proposed to gain accuracy in image classification. CNNs capture features of images with “strong and mostly correct assumptions about the nature of images”

Overall, this study is to identify a single object in an image and to classify it into one of the recycling categories, such as metal, paper, and plastic.

Problem Statement

Humans generated more than 2 billion tons of solid waste in 2016 and by 2050, that could rise to 3.4 billion tons, according to the World Bank. About 12% of all municipal waste in 2016 was plastic, 242 million tons of it. The solution could lie in new technologies and a change in social behavior that reduces and even eliminates the need for landfills and incinerators. Sorting trash can be an unpleasant job, one reason why a lot of rubbish ends up in developing countries with lower wages.

Finding an automatic way to do the recycling is now of great value to an industrial and information-based society, which has not only environmental effects but also beneficial economic effects. Since last decade the industry of artificial intelligence has welcomed its third wave with a sufficient database. Deep learning began to show its high efficiency and low complexity in the field of computer vision.

Overall, this study is to identify a single object in an image and to classify it into one of the recycling categories, such as metal, paper, and plastic.

Dataset And Data Collection

For garbage classification, we utilize the images of the dataset dedicated to the garbage classification. Data set contains various garbage images like cardboard, pieces of plastic, glass, metal cans , etc.

To train deep neural networks, we need a large amount of training images. Dataset contains more than 3 thousand of waste/garbage images to train the CNN model.

Simple CNN Architecture

To investigate performance of a basic CNN, we build a simple CNN architecture to get general inspection, which may help to realize the performance difference between models. This architecture uses 2D convolutional (conv. In short) layers to

capture features of images. Since filters of size 3×3 allow more applications of nonlinear activation functions and decrease the number of parameters than larger filters, the built simple CNN model uses 3×3 filters for all the conv. layers. Between 2D conv. layers we add the max pooling layers to reduce dimensions of the input and the number of parameters to be learned. This could preserve important features after conv. layers while preventing overfitting. After the conv. blocks there is a flatten layer, which flattens the feature matrix into a column vector. This allows the model to use two fully connected layers at the end to do the classification. In this architecture, we use two activation functions. In all the conv. layers and after the flatten layer we use the Rectified Linear Unit function (ReLU) defined as $y = \max(0, x)$ to introduce nonlinearity into the model, which could avoid the problem of gradient vanishing during backpropagation and has a lower calculation complexity. In the last dense layer, we use the softmax function as activation, which fits the cross entropy loss function well.

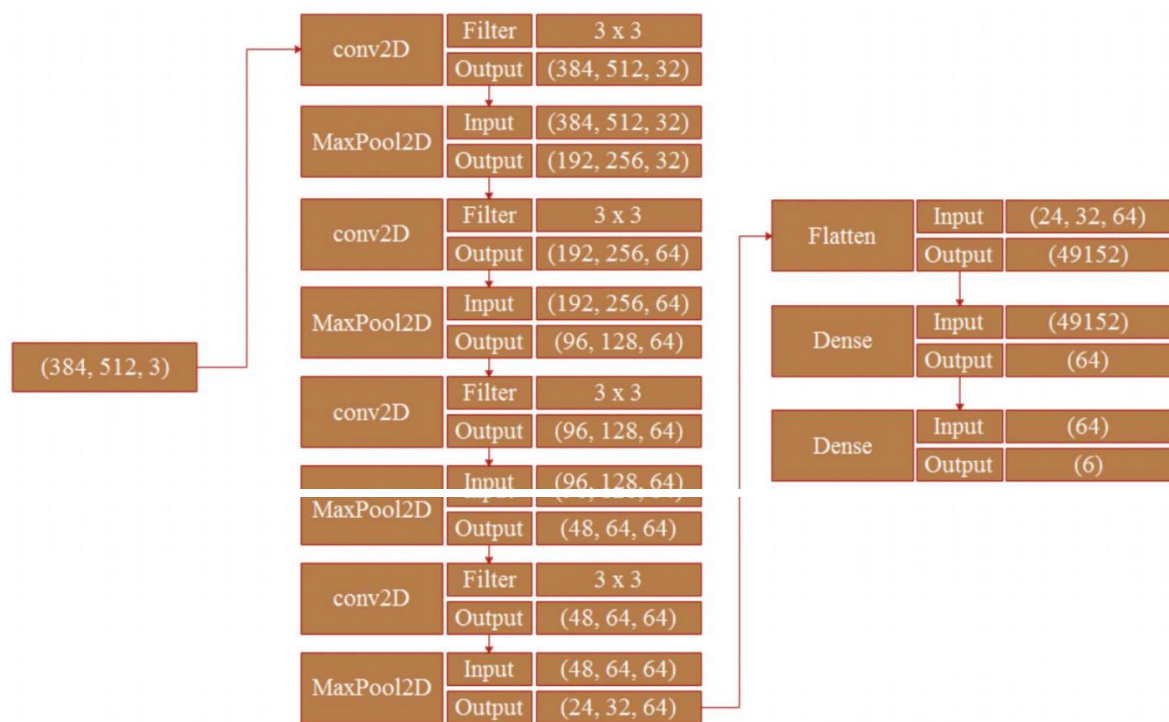


Fig. Structure of simple CNN.

The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision. For example, given many pictures of cats and dogs it learns distinctive features for each class by itself.

Approach

To train the model we have input data and the expected output data.

To create such model, we have to go through the following phases:

- model construction
- model training
- model testing
- model evaluation

Model Construction

From the problem at hand we understand we need to classify the input into either of the six categories of garbage.

So we need to look at the Classification Algorithm.

Since we are working with Images we have to use CNN(Convolutional Neural Networks)for e.g. Keras Sequential model

To start, the CNN receives an input feature map: a three-dimensional matrix where the size of the first two dimensions corresponds to the length and width of the images in pixels. The size of the third dimension is 3 (corresponding to the 3 channels of a color image: red, green, and blue). The CNN comprises a stack of modules, each of which performs three operations.

1. Convolution

A convolution extracts tiles of the input feature map, and applies filters to them to compute new features, producing an output feature map, or convolved feature (which may have a different size and depth than the input feature map).

Convolutions are defined by two parameters:

- Size of the tiles that are extracted (typically 3x3 or 5x5 pixels).
- The depth of the output feature map, which corresponds to the number of filters that are applied.

During a convolution, the filters (matrices the same size as the tile size) effectively

slide over the input feature map's grid horizontally and vertically, one pixel at a time, extracting each corresponding tile

2. ReLU:

Rectified linear unit activation function (or ReLU, for short) often works a little better than a smooth function like the sigmoid, while also being significantly easier to compute. The ReLU function, $F(x)=\max(0,x)$, returns x for all values of $x > 0$, and returns 0 for all values of $x \leq 0$.

3. Pooling

After ReLU comes a pooling step, in which the CNN downsampled the convolved feature (to save on processing time), reducing the number of dimensions of the feature map, while still preserving the most critical feature information. A common algorithm used for this process is called max pooling. Max pooling operates in a similar fashion to convolution. We slide over the feature map and extract tiles of a specified size. For each tile, the maximum value is output to a new feature map, and all other values are discarded. Max pooling operations take two parameters:

Size of the max-pooling filter (typically 2x2 pixels)

Stride: the distance, in pixels, separating each extracted tile. Unlike with convolution, where filters slide over the feature map pixel by pixel, in max pooling, the stride determines the locations where each tile is extracted.

Fully Connected Layers

At the end of a convolutional neural network are one or more fully connected layers (when two layers are "fully connected," every node in the first layer is connected to every node in the second layer). Their job is to perform classification based on the features extracted by the convolutions. Typically, the final fully connected layer contains a softmax activation function, which outputs a probability

value from 0 to 1 for each of the classification labels the model is trying to predict.

Prevent Overfitting: As with any machine learning model, a key concern when training a convolutional neural network is overfitting: a model so tuned to the specifics of the training data that it is unable to generalize to new examples. Two techniques to prevent overfitting when building a CNN are:

Data augmentation: artificially boosting the diversity and number of training examples by performing random transformations to existing images to create a set of new variants. Data augmentation is especially useful when the original training data set is relatively small.

- Dropout regularization: Randomly removing units from the neural network during a training gradient step.

- **Model Training**

Once the Model Layers are finalized we can proceed with Training

First we perform the train test Split to get our training and testing datasets.

Then we can train the model with the training data.

- **Model Testing**

Model Testing can be performed on the split testing dataset. In this case we also have a test dataset separately available that can be used for validation.

- **Model Evaluation**

For evaluation purposes we need to select a baseline model. For this purpose we can use a simple Logistic Regression model to get a baseline score. The metrics to be used will depend on the models being selected and the relevant metric that can be used. Thereafter on implementation of the CNN model we can compare and try to improve on the performance of the CNN model.

Drawbacks

The accuracy of the whole system majorly depends on the dataset used for classifying purposes. As trash can consist of a wide range of items that also appear different depending on how, when and for how long they have been disposed of, the dataset we use has some limitations. To increase the accuracy of our system, more real data of actual piles of garbage are also required to train our model better.