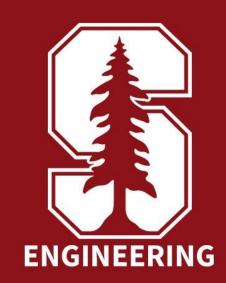# The Price is Right: Predicting Prices with Product Images

Steven Chen, Edward Chou, Richard Yang
{stevenzc, ejchou, rry}@stanford.edu
Department of Computer Science, Stanford University

Motivation

- **Main Goals:**
  - Predict the price of a product given an image
  - Visualize the features that result in perceived higher or lower prices

- **Target Applications:**
  - Help sellers determine what visual features are perceived to be expensive and cheap
  - Assist in valuations of products (eBay, auction sites)
  - Automatically detect inaccurate or unreasonable pricing

Datasets

We collect and clean two novel datasets for this project:
- **Bikes** - bike images and prices, 22,000 examples
  - Prices range from $70 to $17,000
- **Cars** - auto images and prices, 1,500 examples
  - Prices range from $12,000 to $2,000,000
- Use 90/10 train/test splits

Predictive Models

- **Linear Regression:**
  - Two different feature descriptors: histogram of oriented gradients (HOG), CNN features, compressed with PCA
  - Use as baseline for deep neural network

- **Multiclass Support Vector Machine (SVM):**
  - Use compressed CNN features as inputs, with price segments as classes
  - Optimize in the dual form, using RBF (Gaussian) kernel

- **Deep Convolutional Neural Network:**
  - We use transfer learning. Freeze the convolutional layers of the pre-trained VGG16 [13] network
  - Add new ReLU fully connected layer
  - Add either a linear output layer (for regression) or softmax output layer (for price segment classification)

Evaluation

- **Regression:**
  - Metrics: root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2)
  - Include a baseline that always predicts the average
  - Deep CNN significantly outperforms other models in all metrics

- **Classification:**
  - Evaluate SVM and deep CNN on predicting price segment classes (4 classes for bikes, 5 for cars)
  - CNN outperforms SVM on bikes, with similar performance on cars

Model	RMSE	MAE	R^2
Average Baseline	1810.19	1318.53	0.00
LinReg (HOG Features)	1274.99	833.02	0.50
LinReg (CNN Features)	1054.67	712.63	0.66
Deep CNN	**747.42**	**405.50**	**0.83**

Table 1: Regression results for the bikes dataset.

Model	RMSE	MAE	R^2
Average Baseline	76240.41	44410.57	0.00
LinReg (HOG Features)	41898.48	27588.70	0.70
LinReg (CNN Features)	37808.84	23929.67	0.75
Deep CNN	**12363.65**	**7477.74**	**0.97**

Table 2: Regression results for the cars dataset.

Model	Precision	Recall	F1-Score
SVM on bikes	0.80	0.45	0.43
Deep CNN on bikes	0.74	0.75	0.74
SVM on cars	0.83	0.82	0.82
Deep CNN on cars	0.82	0.82	0.82

Table 7: Classification results for the bikes and cars datasets.

Visualization Methods and Results

- **Sliding Window Heatmaps** [16]
  - Cover up blocks of the image to see how it effects the regression output, and color its effects
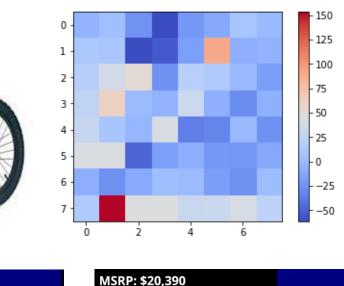 - Obscuring the training wheels greatly increases the predicted price

- **Saliency Maps** [12]
  - Compute gradient of the class score with respect to the input image at every pixel
  - Saliency value is the maximum of the gradient at each pixel across all color channels
  - Features such as the seat and handlebars of the bikes, as well the brand and body contours of the cars are particularly significant

- **Class Activation Maps** [11]
  - Weighted sum of spatial averages of feature maps at each unit at the last conv layer, overlaid on top of original image
  - CAM for bikes confirms significant areas are brakes, gears, seats, wheels
  - CAM for cars highlights the doors (2 vs. 4 door) and open top convertibles

Conclusion

- We apply a variety of techniques for price prediction, using linear regression (w/ PCA), SVM classification, and deep learning
- We show that for both regression and classification, deep CNN performs strongest
- Deep CNN achieves MAE of ~$400 on a ~$17,000 bike price range, and ~$7,500 for a ~$2,000,000 car price range
- Visualizations reveal insightful results; saliency maps and CAMs show most impactful features include the brand logo of the Volkswagen beetle, the open hood of the convertible, and the seats, wheels, and handles of the bikes
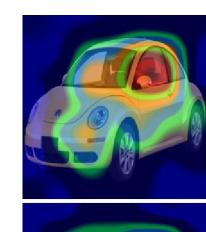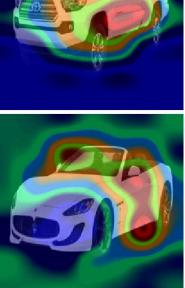
Future Work + Applications

- Work with messier image datasets, containing different angles and backgrounds
- Use generative visualization techniques for different price segments
- Helps designers create products that visually convey expensiveness
- Create model for used cars or bikes, used as an appraisal tool
- Training on other product categories, to automate the selection of a starting bid using a photo for auction sites (eBay, Craigslist)

References

[1] F. Chollet et al. Keras. https://github.com/fchollet/keras, 2015.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005.

[3] Y. Eisenthal. Facial Attractiveness: Beauty and the Machine. PhD thesis, Tel-Aviv University, 2006.

[4] Y. Fu and T. Huang. Human age estimation with regression on discriminative aging manifold. IEEE Transactions on Multimedia, 10(4), 2008.

[5] H. Han, C. Otto, and A. Jain. Age estimation from face images: Human vs. machine performance. In IAPR International Conference on Biometrics, 2013.

[6] R. Kotikalapudi and contributors. keras-vis. https://github.com/raghakot/keras-vis, 2017.

[7] V. Limsombunchai. House price prediction: Hedonic price model vs. artificial neural network. In NZARES, 2004.

[8] A. Maurya. Clicktoprice: Incorporating visual features of product images in price prediction. In INFORMS, 2016.

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[10] A. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition. In CVPR, 2014.

[11] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In ICCV, 2017.

[12] K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. In ICLR, 2014.

[13] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In ICLR, 2015.

[14] J. Yosinski, J. Clune, A. Nguyen, T. Fuchs, and H. Lipson. Understanding neural networks through deep visualization. In ICML, 2015.

[15] Q. You, R. Pang, L. Cao, and J. Luo. Image-based appraisal of real estate properties. IEEE Transactions on Multimedia, 19(12), 2017.

[16] M. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In ECCV, 2014.

[17] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In CVPR, 2016.