Every day, we encounter a large number of images from various sources such as the internet, news articles, document diagrams and advertisements.The main aim of this paper is to provide a comprehensive survey of deep learning for image captioning.In traditional machine learning, hand crafted features such as Local Binary Patterns (LBP) [107], Scale-Invariant Feature Transform (SIFT) [87], the Histogram of Oriented Gradients (HOG) [27], and a combination of such features are widely used.Image indexing is important for Content-Based Image Retrieval (CBIR) and therefore, it can be applied to many areas, including biomedicine, commerce, the military, education, digital libraries, and web searching.For example, Convolutional Neural Networks (CNN) [79] are widely used for feature learning, and a classifier such as Softmax is used for classification.CNN is generally followed by Recurrent Neural Networks (RNN) in order to generate captions.These survey papers mainly discussed template based, retrieval based, and a very few deep learning-based novel image caption generating models.Social media platforms such as Facebook and Twitter can directly generate descriptions from images.Generating well-formed sentences requires both syntactic and semantic understanding of the language [143].Since hand crafted features are task specific, extracting features from a large and diverse set of data is not feasible.On the other hand, in deep machine learning based techniques, features are learned automatically from training data and they can handle a large and diverse set of images and videos.Although the papers have presented a good literature survey of image captioning, they could only cover a few papers on deep learning because the bulk of them was published after the survey papers.To provide an abridged version of the literature, we present a survey mainly focusing on .the deep learning-based papers on image captioning