**The problem statement is to consider incorporating sentiment analysis to generate captions that capture the emotions and mood of the images.**

The solution is to generate image captions that reflect the emotions and mood of the images can add a new dimension to our image captioning system and this can be particularly valuable for applications like social media, photography websites, or any platform where images are shared. Here we can implement by,

**1. Data Collection and labelling:**

Gather a dataset of images along with associated sentiment labels or tags. These labels can range from basic emotions like "happy," "sad," "angry," to more nuanced sentiments. We may need human annotators to label the data.

**2. Sentiment Analysis Model:**

Train or use a pre-trained sentiment analysis model. We can choose from various techniques, such as traditional machine learning models or deep learning models. Fine-tune this model using the labelled data to predict sentiment from text.

**3. Image Feature Extraction:**

Employ a pre-trained Convolutional Neural Network like VGG, ResNet, or Inception to extract high-level visual features from the images. These features should represent the visual content of the images.

**4. Integration of Sentiment and Visual Features:**

Combine the sentiment analysis model and image features into a single feature vector. This could involve concatenating or merging these features, depending on the chosen architecture.

**5. Captioning Model:**

Utilize a caption generation model, such as a Recurrent Neural Network or Transformer based model. This model will take the integrated feature vector as input.

**6. Training the Captioning Model:**

Train the captioning model on a dataset of images paired with both image features and their associated sentiment labels. The captions generated by the model should ideally reflect the mood or emotion described in the sentiment labels.

**7. Caption Generation:**

During inference, provide an image as input to the captioning model. Extract image features using the CNN and predict sentiment using the sentiment analysis model. Merge the sentiment prediction and image features, then feed this merged information to the captioning model to generate a caption that reflects the emotions and mood of the image.

**8. Post-processing and Fine-tuning:**

After generating captions, apply post-processing to ensure coherence and relevance. We may also fine-tune the model based on user feedback and continuous evaluation.

**9. Evaluation:**

Assess the quality of the generated captions using human evaluators or automated metrics. Evaluate whether the captions effectively capture the emotions and mood of the images.

**10. Deployment:**

Integrate the sentiment-aware image captioning system into web application, website, or platform. Ensure a user-friendly interface and seamless user experience.

**11. User Interaction:**

Allow users to provide feedback on the generated captions, which can help in improving the model over time and making it more aligned with users' preferences.

By following this design, we can create a system that combines visual content analysis with sentiment analysis to generate image captions that are not only descriptive but also emotionally resonant, providing a richer experience for users who engage with images on our platform.