Fig. 7. Expression level evaluation results of S_1 (top left), S_2 (top right) and S_3 (bottom left). The solid lines represent origin results and the dashed ones represent results with less symbol evaluated.
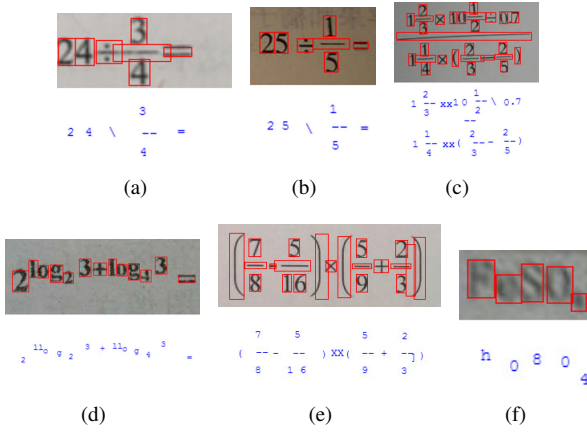

Fig. 8. Results of context-aware framework for hard tackling cases. (a) Auto-cut touching symbols. We use '‑‑' to represent fraction line and \ to represent '÷'; (b) (c) Auto-combine multi-parts symbols like '÷' '='. 'xx' that represents multiplication sign is used to distinguish from letter 'x'; (d) Ambiguous symbol '1', 'o' in 'log' is correctly recognized. We use 'll' here to distinguish from digit '1'; (e) (f) Failed cases for long symbols like brackets and blurred ones

Fig. 9. (a) The structure without shared convolutional features. (b) The structure with no detection. (c) The proposed method

C. Case Analysis

The purpose of context-aware framework is to solve the existing problems in ME recognition. The Fig 8 displays some hard-tackling cases for previous methods and results of our model. The results in Fig 8 display the effectiveness of proposed method to handle multi-parts symbols like '÷', ':', touching case in Fig 8.a and ambiguous symbols like fraction line, minus sign, letter 'l' and digit '1'. We also display some failed cases like long and blurred symbols. For long symbols like bracket, regression task tends to output poor results and for blurred images like Fig 8.f, we human could recognize the symbols '$FeSO_4$' depending on our much more abundant prior knowledge.

D. Effectiveness of Multi-task Learning

To test the effectiveness of triple tasks, we have designed two extra networks. The first one removes detection task and takes the backgrou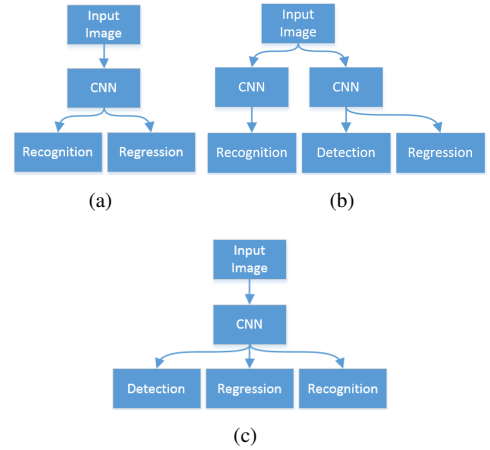nd category into recognition. This no-detection network is designed to demonstrate the importance of detection task though the recognition task could also handle detection. The second network trains recognition task separately with the same convolution structure as detection and regression tasks. The latter network intends to prove the benefit of shared feature design. Simplified structures of three networks for comparison are shown in Fig 9.

The results are shown in Table 4. By comparison we can come to the conclusion that triple-task structure indeed benefits the performance. The no-detection network gives a much lower precision which indicates the idea that splitting a complicated task into two indeed helps improve the overall performance. For the unshared feature structure, precision and recall are both slightly lower than that of the proposed method. It is obvious that the shared feature structure could provide enough information and save much more computation than the other.

VI. CONCLUSION

In this paper, we present a context-aware end-to-end system for ME recognition, by introducing a CNN based structure with multi-task learning to perform mathematical symbols detection and recognition simultaneously. Experiments verify